CYANOBACTERIA-ASSOCIATED BACTERIOPHAGE COMMUNITIES ACROSS SCALES OF SPATIAL, TEMPORAL AND ENVIRONMENTAL CHANGE

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Julia Margaret Brown

January, 2015

CYANOBACTERIA-ASSOCIATED BACTERIOPHAGE COMMUNITIES ACROSS SCALES OF SPATIAL, TEMPORAL AND ENVIRONMENTAL CHANGE

Julia Margaret Brown, Ph. D.

Cornell University 2015

Viruses are prevalent in most, if not all, environments on Earth. They contribute significantly to global genetic diversity and interact with organisms in all branches of life.

Although the field of viral ecology has made great strides in understanding role of viruses in ecosystems functions, the complexity and diversity of viruses have made them particularly difficult to characterize. This dissertation contributes to the overall understanding of viral ecology through investigation into bacteriophage responses to varying scales of change utilizing unique characteristics of two different environments. Chapter 1 introduces key concepts in viral ecology that will be addressed in the body of the dissertation. Chapter 2 assesses virus-host dynamics *in situ* by following the dynamics of three cyanobacterial viruses and the cyanobacterial population over two summer seasons in Fayetteville Green Lake and Round Lake, neighboring lakes that are biogeochemically similar. This study found consistent trends in cyanophage abundance between lakes, but annual variability and a weak relationship between the abundance of virus and host populations. Chapter 3 focuses on the marine environment to examine how a viral community responds when exposed to sharp environmental change in the form of cyanobacterial aggregation collapse. This study observed both lytic and lysogenic phages of bacteria known to be associated with *Trichodesmium* colonies, showing that viruses capable of both infection strategies respond to environmental change. Chapter 4 returns to Fayetteville Green Lake to probe temporal aspects of virus population dynamics on extended timescales. This chapter examines signatures of

cyanobacterial viruses in over 180+ years of the lake's history through examination of signatures of cyanobacterial virus genes present within chronologically distinguishable, undisturbed sediment cores. This study found patterns of representation of genetic signatures of cyanophages that suggest a dynamic population of cyanophages that have varying evolutionary trajectories within the lake. Through providing insights into how viral communities change over several scales of inquiry, the work presented in this dissertation adds to the growing body of knowledge of viruses in the environment and offers unique insights that will greatly contribute to the field of viral ecology.

BIOGRAPHICAL SKETCH

Julia Margaret Brown was born in Danbury, Connecticut on January 15th 1986 to Nancy and Frank Brown. She grew up in Weston, Connecticut with her sisters Becky and Molly and spent the summers of her childhood at a family cottage on the coast of Maine. She attended Weston Public Schools for grades K through 12. She graduated from Weston High School in 2004.

In 2008, she received a Bachelor of Arts degree from Carleton College in Northfield, Minnesota where she majored in Chemistry with a concentration in Biochemistry. Julie first learned about marine viruses during her senior year at Carleton while pursuing off campus study with the Sea Education Association in Woods Hole, Massachusetts. While designing an independent research project for her time at sea, she sought to find an area that combined her interests in environmental science, biochemistry and chemistry. After being informed that there were millions of viruses and bacteria in a drop of seawater, Julie sat for hours in the old annex of the Marine Biological Laboratory Library pouring through primary literature about microbial oceanography. Having found an area that she felt matched her interests, Julie applied to the microbiology graduate program at Cornell University to pursue graduate research in environmental microbiology.

Julie began graduate study at Cornell University in the fall of 2008. On the first day of graduate student orientation, Julie sat down next to another first year student in the microbiology graduate program named Kevin who quickly became her best friend. Julie married Kevin Posman six years later on the shore of a nearby lake.

Julie's research while at Cornell in the lab of Dr. Ian Hewson explored the dynamic nature of viral populations in aquatic environments *in situ* using a variety of molecular and bioinformatic techniques.

To Kevin, without whom none of this would've been any fun at all. I love you.

To my parents, thank you for all you've given me. I'll do my best to pay it forward.

ACKNOWLEDGMENTS

First and foremost, I would like to thank Dr. Ian Hewson for his guidance and support during my time at Cornell. I have learned many things and grown as a scientist under his direction. I am thankful for the many opportunities he has given me during the last five years. This research is made possible through his support.

I would like to thank my committee members, Dr. Nelson Hairston and Dr. Daniel Buckley for their advice and guidance. There were points where I felt lost during this process. Meeting with both of my committee members one-on-one and during committee meetings always left me feeling invigorated and excited about my project.

Thank you to the many faces in the department of Microbiology who have made my time at Cornell a truly enriching experience. Thank you to Sue Merkel and the teaching staff for showing me how to teach students about the exciting world of microbes, and Dr. Esther Angert and Dr. Stephen Zinder for leading me as I assisted the Advanced Laboratory in Microbiology. I would like to acknowledge Shirley Cramer, Pati Butler, Cathy Shappell, Lillian Henry and Michelle Carr for their administrative support while I was at Cornell. You ladies are amazing.

Thank you to the following people for assisting me with field sample collection during various aspects of the Green Lake work: John McAdams, Phil Thompson, Brent Gudenkauf, James Eaglesham, Dr. Cresten Mansfeldt, Kevin Posman, Erin Eggleston, Elliot Jackson, Dr. Buck Hansen, Dr. Madeline Galac, Dr. Heather Fullerton, Andrew Myers, Nancy Brown and Frank Brown. Kevin and Cresten were instrumental in sediment sample collection. Without them, I would have given up after two unsuccessful attempts. Sample transport was made possible by a 1996 Toyota Camry and the little red wagon. Thank-you to Tom Hughes, Jim Semar and the staff at Green Lakes State Park. Samples in GL and RL were collected under New York State Park permits 2012-GR-001, 2012-GR-007, 2013-GR-001 and 2014-GR-003.

Dr. Brooks Miner and Dr. Cresten provided statistics advice for the time series chapter.

Kevin Posman, Cresten Mansfeldt, Erin Eggleston and Armanda Roco proofread different parts of this dissertation. Sheila Saia assisted in SRP measurement and Lindsay Schaffner provided support and advice for chlorophyll a measurement. Thank you to Dr. Nelson Hairston for lending his gravity corer for sediment core collection. For the *Trichodesmium* project, Ian Hewson conducted the fieldwork and sample collection and Brenna LaBarre provided bioinformatics assistance.

Thanks to the many friends in Ithaca who have made being a graduate student at Cornell a fun and enriching experience. Thank you to the Ithaca Ultimate Frisbee community who immediately welcomed me into the Ithaca Ultimate scene. I have had a blast playing on many teams including 7 recreational teams, FLOAT, Vitamin I and the Townies. Thank-you to the lady's night girls, the Bowlologists, the innertube water polo team and many other friends who have made the past six years a truly great experience. I would particularly like to acknowledge Mario Martone, Erin Eggleston, Ashley Campbell, Armanda Rocco and Cresten Mansfeldt for their friendship throughout my time in Ithaca.

Thank you to my family for their love, support and enthusiasm. Thank you to the Posmans for taking care of me during my writing phase while Kevin was on his grand adventure. Thank you to my parents for being my number one fans and for supporting me in more ways than can be written on a page. Thank you to my sisters for being the friends I will never lose touch with no matter how busy life gets; who show me interesting and different perspectives on the world.

Finally, I would like to thank my husband, Kevin Posman. Thank you for being my stress relief, for bringing me back to earth when things get crazy. I appreciated your patience as I finished up my research and dissertation. Your love and support is all encompassing, even when we are a sea apart. I am so lucky to have you in my life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1. Introduction

1.1 The importance of viruses in the environment

Viruses are prevalent in virtually all environments on Earth. They contribute significantly to global genetic diversity and interact with organisms in all branches of life. Viruses represent an abundant and dynamic component of the microbial community from soils (Fierer et al., 2007), to sediments (Breitbart et al., 2004; Hewson et al., 2003), to reclaimed and potable water (Rosario et al., 2009), aquatic habitats (reviewed in Hewson et al., 2010), to the human microbiome (reviewed in Wylie et al., 2012), to extreme environments (e.g. Breitbart et al., 2004; Emerson et al., 2012; Heidelberg et al., 2009; Sime-Ngando et al., 2011; Williamson et al., 2008). The high abundance of viruses in the environment was first discovered in marine habitats (Bergh et al., 1989; Proctor and Fuhrman, 1990), and since then, aquatic environments have been a central area of study to understand the dynamics of viral communities. Concentrations of virus-like particles range from 10^8 to 10^{11} per milliliter in aquatic environments (Weinbauer, 2004). A large portion of viruses within aquatic habitats are known to be double stranded DNA viruses that infect co-occurring bacteria (Weinbauer, 2004; Wommack and Colwell, 2000). Viruses influence bacterial communities and ecosystem processes in a variety of ways. They shape microbial community structure through host mortality, which in turn contributes to carbon cycling and influence community level metabolic processes (Fuhrman, 1999; Suttle, 2007; Wommack and Colwell, 2000). Viruses contribute to bacterial genetic diversity on an individual level through co-evolutionary interactions with host species (Marston et al., 2012; Rodriguez-Valera et al., 2009; Thingstad et al., 2014) and at a community scale by controlling the abundance of community constituents through "kill-the-winner" infection dynamics (Thingstad, 2000; Thingstad and Lignell, 1997; Winter et al., 2010). Viruses also contribute to host metabolism by encoding auxiliary metabolic genes that are expressed during infection (Clokie and Mann, 2006; Lindell et al., 2007). Finally, the viral community as a whole represents a significant pool of genetic material, which contributes to the

overall genetic diversity of the environment. Viral metagenomes contain a vast amount of as yet uncharacterized genetic information, arguably the most unknown genetic diversity on earth (Breitbart, 2012). Not only are viruses diverse, but their populations are dynamic; the structure of these viral consortia are constantly changing (Rodriguez-Brito et al., 2010).

With such significant influences on host diversity, metabolism, and genetic potential within an environment, understanding the factors that shape viral diversity is critical. This dissertation seeks to understand the mechanisms of change within viral communities from three different perspectives: through seasonal variations in virus-host dynamics, through changes in viral communities in response to sudden environmental change and through viral communities over longer periods of time within the same location. Each investigation focuses on comparative analyses of subsets of the viral population across differing temporal, spatial and environmental scales to address how these specific factors contribute to viral community structure.

1.2 Virus-host relationships in aquatic environments

Many evolutionary and physical factors influence *in situ* relationships between viruses and hosts. Investigations of the dynamics between a single virus and its host in a controlled, laboratory setting highlights the complexity of this interaction. In laboratory experiments, virus-host dynamics are observed to be predictable (Middelboe, 2000) but the nature of the relationships vary by virus-host pair (Moebus, 1996a) and between different experimental conditions (Moebus, 1996b). In addition, co-evolutionary interactions between viruses and hosts in continuous batch culture leads to the persistence of resistant bacterial populations that co-exist with viral populations (Middelboe et al., 2001); through interacting with a virus, the relationships between a host population and a virus changes.

Studies of *in situ* dynamics between viruses and hosts examine community level patterns in bacterial and viral populations rather than exploring individual relationships. Such investigations

result in mixed success for observing predictable relationships between viral and bacterial communities. Some studies were able to find a relationship between bacterial and viral population structure and diversity (Chow et al., 2013; Hardbower et al., 2012; Hewson et al., 2006; Mühling et al., 2005; Needham et al., 2013; Sandaa and Larsen, 2006) while others were not (Lymer et al., 2008; Tijdens et al., 2008; Zhong et al., 2013). Many of the investigations that observed relationships between viral and bacterial populations were conducted in the marine environment, where more extensive investigations into viral contributions have been conducted and where dynamics of host populations are seasonally predictable (Fuhrman et al., 2006). Although host communities have been shown to follow predictable seasonal patterns in some freshwater lakes (Nelson, 2008), freshwater host populations can be variable between years (Villar-Argaiz et al., 2001), and the limited success of cross-lake comparisons lead to varying observations of factors influencing the abundance of viruses and hosts through time(Lymer et al., 2008; Zhong et al., 2013). Large scale, high resolution studies over longer time periods have had more success observing trends between bacterial and viral populations (Chow et al., 2013), as have studies that employed the same high resolution method over short periods of time (Needham et al., 2013). While these marine studies have had success identifying relationships between virus and host populations, no studies in lacustrine systems have described a consistent and predictable relationship between viruses and their bacterial host populations across lakes, and few have found a relationship between the two populations within the same lake (Hardbower et al., 2012; Lymer et al., 2008). Although the co-occurrence of viruses and hosts appears to be predictable in the marine environment, whether such patterns are observed across aquatic environments has not been established.

1.3 Viral community responses to environmental change

Understanding how communities shift in the face of environmental change is a fundamental ecological question. Viral communities are dynamic, even when presented with seemingly static

environmental conditions (Rodriguez-Brito et al., 2010). It follows then that drastic changes within the environment would result in significant changes in viral community structure. Shifting environmental conditions, such as a sudden addition of organic matter to an otherwise oligotrophic system leads to changes in bacterial metabolism and subsequent changes in bacterial assemblage structure (Carlson et al., 2002). These changes within the host population likely lead to a response from the corresponding phage population, whose infection dynamics change in differing nutrient conditions (Middelboe, 2000; Moebus, 1996b).

In addition, a shift in host metabolism due to changes in the environment may lead to induction of prophage into a lytic infection cycle (Williamson et al., 2002). Prophage are lysogenic viruses that have integrated into the host's genome and lie dormant until some environmental trigger leads to the induction of the prophage from a lysogenic to a lytic state (Weinbauer, 2004). Exactly what induces a prophage to transition to lytic infection in the environment is not entirely established, but prophage induction may be dependent on host density, with lysogeny being more prevalent in times of low host abundance (McDaniel and Paul, 2005; McDaniel et al., 2002). As such, latent phage of bacteria that respond most strongly to sudden environmental change would be expected to become prevalent within the viral community after the change has occurred.

Nutrient addition studies mimic sharp environmental changes and offer insights into the viral response to such events. Studies monitoring viral abundance in response to nutrient addition report increases in viral abundance in nutrient amended incubations (Hewson et al., 2003; Ram and Sime-Ngando, 2008) as well as changes in viral community structure (Ovreas et al., 2003; Sandaa and Larsen, 2006).

Phytoplankton blooms are an example of an aquatic environmental setting in which bacterioplankton and virioplankton populations experience dramatic changes in environmental conditions over short time periods. Such events lead to a change in the nature of organic matter available to co-occurring microorganisms at bloom onset, maturation and collapse. A study of

bacteria and viruses associated with an arctic algal bloom showed an increase in the abundance of viruses relative to bacteria in the later stages of the bloom, indicating enhanced viral activity (Yager et al., 2001). This relationship was also observed when investigating bacterial and viral populations associated with *Karenia brevis* blooms in the Gulf of Mexico (Meyer et al., 2014).

The chemical mitomycinC (mitC) is used to directly address the presence and contribution of lysogenic populations in mesocosm experiments. MitC is a antibiotic that induces bacterial SOS response via crosslinking of DNA (Iyer and Szybalski, 1963; Tomasz, 1995). MitC is used in viral ecology as chemical inducing agent that has been shown to induce transition of prophages to a lytic cycle within natural microbial communities (Weinbauer, 2004). As mitC stimulates lytic production in viruses integrated into host genomes, it leads to changes in the viral population that provides insight into the prevalence of lysogeny within a host population, and introduces an entirely different population of viruses into the planktonic phase (McDaniel et al., 2008, 2014; Williamson et al., 2008). For example, two studies in the Gulf of Mexico examining communities incubated with and without mitC showed that mitC treated water revealed a different population of viruses than the untreated incubation (McDaniel et al., 2008, 2014), thus showing the potential for inducing conditions to drive differences in the structure of the active virus community.

Although early studies examined how the viral community responds to the addition of nutrients and organic matter (Hewson et al., 2003; Ovreas et al., 2003; Ram and Sime-Ngando, 2008), little is known about how such change influences the infecton strategies and overall structure of the viral community. A study in deep sea hydrothermal vents showed that there is a higher prevalence of lysogeny in vent water compared to surrounding seawater, suggesting that stressful conditions for the host population prevents lytic conversion of lysogenic phages (Williamson et al., 2008), but no converse observation has been made about community responses to influxes of organic matter that would presumably promote bacterial production. Although there is a lack of metagenomic information characterizing how viral communities respond to

environmental change, understanding how viral populations with differing infection strategies respond to rapid change is a concept that may be applied across environments.

1.4 Viral community dynamics through time

Several investigations have found that on the timescale of days to years, the diversity of viruses through space and time appears constant but community structure is dynamic (Emerson et al., 2012, 2013; Rodriguez-Brito et al., 2010). Some research observed that viral population structure over short timescales is observed to be tightly coupled with host population dynamics (Needham et al., 2013), and one study showed that the presence of viral populations can be predicted based on the prevalence of host populations (Chow et al., 2013), however diversifying forces such as antagonistic co-evolution and migration make it likely that phage populations shift less predictably when examined on longer timescales. An investigation that characterized the dominant populations of the marine eukaryotic phytoplankton, *E. huxleyi,* and co-occurring viruses over seven millennia found that the dominant virus and host genotypes changed through time, but also observed the recurrence of the same host population centuries apart with a different co-occurring viral type at different points in history (Coolen, 2011). This suggests that through geological time host abundance alone does not determine the composition or structure of the virus population. The variability in viral genotypes through time may be driven in part by the dispersal of viruses between habitats. A study examining the diversity of archaeal virus types within disconnected Yellowstone National Park hot springs found that diversity of these populations was not dispersal limited, suggesting that viruses moved between environments (Snyder et al., 2010).

Studies probing historical changes in viral populations often do so through examination of viral elements within bacterial genomes, such as Clustered Regularly Spaced Palindromic Repeat (CRISPR) regions of bacterial and archaeal genomes (Heidelberg et al., 2009; Tyson and Banfield, 2008). CRISPRs, found in many archaeal, and some bacterial genomes, contain variable insert DNA sequences with homology to viral sequences (Barrangou et al., 2007; Sorek et al., 2008). These

sequences are used by the host, along with a cas gene, as an adaptive immune system to combat viruses present within the environment (Horvath and Barrangou, 2010). Because of this, the CRISPR spacer sequences represent portions of viruses and other mobile genetic elements that the microbe has previously encountered. CRISPR loci have a temporal aspect to them, with the most recently added spacers being closest to the cas gene, while the most ancient sequences falling further downstream from the cas gene (Horvath et al., 2008).

Genomics and metagenomics coupled with identification of CRISPR regions of microbial genomes shows how viruses may change via co-evolution with their hosts (Heidelberg et al., 2009). One study identified CRISPR sequences within isolate genomes and microbial metagenomes of microorganisms from Yellowstone Hot Springs. Comparison of viral metagenomes and CRISPR spacers showed nucleotide-level changes in the viral metagenomes that suggested evolution of the phage genes to evade detection by the CRISPR-cas system (Heidelberg et al., 2009). In acid mine drainage, CRISPR spacers within euryarchaeal genomes assembled from metagenomic sequence data showed shared spacer sequences occupied the most ancient portions of the CRISPR regions and recently acquired spacer sequences were the most uncommon (Tyson and Banfield, 2008). A subsequent study found that only the most recently acquired CRISPR spacer sequences matched sequences within the simultaneously sequenced viral metagenomes, while no viral metagenomic sequences matched the oldest portions of the CRISPR regions (Andersson and Banfield, 2008; Weinberger et al., 2012). Together these results suggest that viral populations are rapidly evolving with host populations, but also that there are substantial shifts in the viral population through time indicated by the absence of viral genotypes found within the CRISPR record from the current population of viruses.

CRISPR studies offer important insights into host interactions with previous viral populations and indicate that virus and host populations are rapidly evolving (Tyson and Banfield, 2008) but they lack temporal clarity. To date, there is no way to determine the point at which the

host population encountered the virus recorded within the host genome, nor do CRISPR sequences reveal information about the total viral community. Finally, CRISPR spacer sequences are very short and offer limited information about the corresponding viral genome from which it originated. Studies of CRISPRs and viral populations are contingent upon obtaining complete sequences of CRISPR regions and adequate sequence coverage of the viral population. These criteria are only possible with either cultivatable virus-host systems and genome sequences of both the virus and host populations, or high sequence coverage of the viral and microbial metagenomes. As the majority of microbial species cannot be cultivated using current techniques, and viral and microbial assemblages are often so diverse that adequate sequence assembly is difficult, CRISPR studies have been most successful in characterizing population dynamics in extreme low diversity environments. To understand virus-host interactions, and how viral populations change through time in more diverse populations, higher sequence coverage and defined temporal estimates are necessary.

1.5 In this dissertation

The following chapters examine changes in cyanobacteria-associated bacteriophage communities from three different perspectives utilizing particular characteristics of two different environments. Chapter 2 assesses virus-host dynamics *in situ* by following the dynamics of three cyanobacterial viruses and the cyanobacterial population over two summer seasons. This study takes place in two biogeochemically similar lakes in an attempt to remove variability seen in studies of other lacustrine systems thus addressing the hypothesis that virus-host dynamics will be similar when lakes are nearly identical. Chapter 3 focuses on the marine environment to examine how a viral community responds when exposed to sharp environmental change in the form of cyanobacterial aggregation collapse, and addresses the hypothesis that sudden environmental change will lead to a shift in the structure of the viral community driven by prophage induction.

Chapter 4 returns to the lacustrine environment to probe temporal aspects of virus population dynamics on extended timescales. This chapter examines signatures of cyanobacterial viruses in over 180+ years of the lake's history through examination of signatures of cyanobacterial virus genes present within chronologically distinguishable slices of undisturbed sediment cores. Through examination of cyanophage diversity and community structure, this chapter addresses the hypothesis that the diversity and structure of cyanophage communities observed in the sediments will exhibit age-dependent relationships, and changes in population structure will be influenced by both mutation and migration on this extended timescale.

No matter what the scale, it is evident that viral community interactions with host populations and the environment are complex. Co-evolutionary interactions, dispersal, unseen environmental factors, variability and the potential for the induction of completely new viral populations complicate our understanding of viral community dynamics on all levels. These chapters utilize a number of different molecular techniques, sequencing technologies and comparative analyses to describe the *in situ* diversity of subsets of the bacteriophage population. Through taking advantage of the characteristics of each environment to address a specific question, this work will contribute to the knowledge of viral community dynamics by characterizing mechanisms that influence change in bacteriophage community structure over varying spatial, temporal and environmental scales.

1.6 References

Andersson, A.F., and Banfield, J.F. (2008). Virus population dynamics and acquired virus resistance in natural microbial communities. Science *320*, 1047–1050.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709–1712.

Bergh, Ø., BØrsheim, K.Y., Bratbak, G., and Heldal, M. (1989). High abundance of viruses found in aquatic environments. Nature *340*, 467–468.

Breitbart, M. (2012). Marine viruses: truth or dare. Mar. Sci. *4*.

Breitbart, M., Felts, B., Kelley, S., Mahaffy, J.M., Nulton, J., Salamon, P., and Rohwer, F. (2004). Diversity and population structure of a near–shore marine–sediment viral community. Proc. R. Soc. Lond. B Biol. Sci. *271*, 565–574.

Carlson, C.A., Giovannoni, S.J., Hansell, D.A., Goldberg, S.J., Parsons, R., Otero, M.P., Vergin, K., and Wheeler, B.R. (2002). Effect of nutrient amendments on bacterioplankton production, community structure, and DOC utilization in the northwestern Sargasso Sea. Aquat. Microb. Ecol. *30*, 19–36.

Chow, C.-E.T., Kim, D.Y., Sachdeva, R., Caron, D.A., and Fuhrman, J.A. (2013). Top-down controls on bacterial community structure: microbial network analysis of bacteria, T4-like viruses and protists. ISME J. *8*, 816–829.

Clokie, M.R., and Mann, N.H. (2006). Marine cyanophages and light. Environ. Microbiol. *8*, 2074–2082.

Coolen, M.J. (2011). 7000 years of Emiliania huxleyi viruses in the Black Sea. Science *333*, 451–452.

Emerson, J.B., Thomas, B.C., Andrade, K., Allen, E.E., Heidelberg, K.B., and Banfield, J.F. (2012). Dynamic viral populations in hypersaline systems as revealed by metagenomic assembly. Appl. Environ. Microbiol. *78*, 6309–6320.

Emerson, J.B., Thomas, B.C., Andrade, K., Heidelberg, K.B., and Banfield, J.F. (2013). New approaches indicate constant viral diversity despite shifts in assemblage structure in an Australian hypersaline lake. Appl. Environ. Microbiol. *79*, 6755–6764.

Fierer, N., Breitbart, M., Nulton, J., Salamon, P., Lozupone, C., Jones, R., Robeson, M., Edwards, R.A., Felts, B., and Rayhawk, S. (2007). Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. Appl. Environ. Microbiol. *73*, 7059–7066.

Fuhrman, J.A. (1999). Marine viruses and their biogeochemical and ecological effects. Nature *399*, 541–548.

Fuhrman, J.A., Hewson, I., Schwalbach, M.S., Steele, J.A., Brown, M.V., and Naeem, S. (2006). Annually reoccurring bacterial communities are predictable from ocean conditions. Proc. Natl. Acad. Sci. *103*, 13104–13109.

Hardbower, D.M., Dolman, J.L., Glasner, D.R., Kendra, J.A., and Williamson, K.E. (2012). Optimization of viral profiling approaches reveals strong links between viral and bacterial communities in a eutrophic freshwater lake. Aquat. Microb. Ecol. *67*, 59–76.

Heidelberg, J.F., Nelson, W.C., Schoenfeld, T., and Bhaya, D. (2009). Germ warfare in a microbial mat community: CRISPRs provide insights into the co-evolution of host and viral genomes. PLoS One *4*, e4169.

Hewson, I., Vargo, G., and Fuhrman, J. (2003). Bacterial diversity in shallow oligotrophic marine benthos and overlying waters: effects of virus infection, containment, and nutrient enrichment. Microb. Ecol. *46*, 322–336.

Hewson, I., Winget, D.M., Williamson, K.E., Fuhrman, J.A., and Wommack, K.E. (2006). Viral and bacterial assemblage covariance in oligotrophic waters of the West Florida Shelf (Gulf of Mexico). J. Mar. Biol. Assoc. U. K. *86*, 591–603.

Hewson, I., Chow, C., and Fuhrman, J.A. (2010). Ecological role of viruses in aquatic ecosystems. eLS.

Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. Science *327*, 167–170.

Horvath, P., Romero, D.A., Coûté-Monvoisin, A.-C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., and Barrangou, R. (2008). Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. J. Bacteriol. *190*, 1401–1412.

Iyer, V., and Szybalski, W. (1963). A molecular mechanism of mitomycin action: linking of complementary DNA strands. Proc. Natl. Acad. Sci. U. S. A. *50*, 355.

Lindell, D., Jaffe, J.D., Coleman, M.L., Futschik, M.E., Axmann, I.M., Rector, T., Kettler, G., Sullivan, M.B., Steen, R., and Hess, W.R. (2007). Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. Nature *449*, 83–86.

Lymer, D., Logue, J.B., Brussaard, C.P., BAUDOUX, A., Vrede, K., and LINDSTRÖM, E.S. (2008). Temporal variation in freshwater viral and bacterial community composition. Freshw. Biol. *53*, 1163–1175.

Marston, M.F., Pierciey, F.J., Shepard, A., Gearin, G., Qi, J., Yandava, C., Schuster, S.C., Henn, M.R., and Martiny, J.B. (2012). Rapid diversification of coevolving marine Synechococcus and a virus. Proc. Natl. Acad. Sci. *109*, 4544–4549.

McDaniel, L., and Paul, J. (2005). Effect of nutrient addition and environmental factors on prophage induction in natural populations of marine Synechococcus species. Appl. Environ. Microbiol. *71*, 842–850.

McDaniel, L., Houchin, L., Williamson, S., and Paul, J. (2002). Plankton blooms: Lysogeny in marine Synechococcus. Nature *415*, 496–496.

McDaniel, L., Breitbart, M., Mobberley, J., Long, A., Haynes, M., Rohwer, F., and Paul, J.H. (2008). Metagenomic analysis of lysogeny in Tampa Bay: implications for prophage gene expression. PLoS One *3*, e3263.

McDaniel, L.D., Rosario, K., Breitbart, M., and Paul, J.H. (2014). Comparative metagenomics: Natural populations of induced prophages demonstrate highly unique, lower diversity viral sequences. Environ. Microbiol. *16*, 570–585.

Meyer, K.A., O'Neil, J.M., Hitchcock, G.L., and Heil, C.A. (2014). Microbial production along the West Florida Shelf: Responses of bacteria and viruses to the presence and phase of< i> Karenia brevis blooms. Harmful Algae.

Middelboe, M. (2000). Bacterial growth rate and marine virus–host dynamics. Microb. Ecol. *40*, 114–124.

Middelboe, M., Hagström, A., Blackburn, N., Sinn, B., Fischer, U., Borch, N., Pinhassi, J., Simu, K., and Lorenz, M. (2001). Effects of bacteriophages on the population dynamics of four strains of pelagic marine bacteria. Microb. Ecol. *42*, 395–406.

Moebus, K. (1996a). Marine bacteriophage reproduction under nutrient-limited growth of host bacteria. I. Investigations with six phage-host systems. Mar. Ecol. Prog. Ser. Oldendorf *14*, 1–12.

Moebus, K. (1996b). Marine bacteriophage reproduction under nutrient-limited growth of host bacteria. II. Investigations with phage-host system [H3: H3/1]. Mar. Ecol. Prog. Ser. Oldendorf *14*, 13–22.

Mühling, M., Fuller, N.J., Millard, A., Somerfield, P.J., Marie, D., Wilson, W.H., Scanlan, D.J., Post, A.F., Joint, I., and Mann, N.H. (2005). Genetic diversity of marine Synechococcus and co-occurring cyanophage communities: evidence for viral control of phytoplankton. Environ. Microbiol. *7*, 499–508.

Needham, D.M., Chow, C.-E.T., Cram, J.A., Sachdeva, R., Parada, A., and Fuhrman, J.A. (2013). Short-term observations of marine bacterial and viral communities: patterns, connections and resilience. ISME J. *7*, 1274–1285.

Nelson, C.E. (2008). Phenology of high-elevation pelagic bacteria: the roles of meteorologic variability, catchment inputs and thermal stratification in structuring communities. ISME J. *3*, 13–30.

Ovreas, L.L., Bourne, D.D., Sandaa, R.R., Casamayor, E., Carrion, S., Goddard, V.V., Smerdon, G.G., Heldal, M.M., and Thingstad, T. (2003). Response of bacterial and viral communities to nutrient manipulations in seawater mesocosms. Aquat. Microb. Ecol.-Pages 31 109-121.

Proctor, L.M., and Fuhrman, J.A. (1990). Viral mortality of marine bacteria and cyanobacteria.

Ram, A.S.P., and Sime-Ngando, T. (2008). Functional responses of prokaryotes and viruses to grazer effects and nutrient additions in freshwater microcosms. ISME J. *2*, 498–509.

Rodriguez-Brito, B., Li, L., Wegley, L., Furlan, M., Angly, F., Breitbart, M., Buchanan, J., Desnues, C., Dinsdale, E., Edwards, R., et al. (2010). Viral and microbial community dynamics in four aquatic environments. ISME J *4*, 739–751.

Rodriguez-Valera, F., Martin-Cuadrado, A.-B., Rodriguez-Brito, B., Pašić, L., Thingstad, T.F., Rohwer, F., and Mira, A. (2009). Explaining microbial population genomics through phage predation. Nat. Rev. Microbiol. *7*, 828–836.

Rosario, K., Nilsson, C., Lim, Y.W., Ruan, Y., and Breitbart, M. (2009). Metagenomic analysis of viruses in reclaimed water. Environ. Microbiol. *11*, 2806–2820.

Sandaa, R.-A., and Larsen, A. (2006). Seasonal variations in virus-host populations in Norwegian coastal waters: focusing on the cyanophage community infecting marine Synechococcus spp. Appl. Environ. Microbiol. *72*, 4610–4618.

Sime-Ngando, T., Lucas, S., Robin, A., Tucker, K.P., Colombet, J., Bettarel, Y., Desmond, E., Gribaldo, S., Forterre, P., and Breitbart, M. (2011). Diversity of virus–host systems in hypersaline Lake Retba, Senegal. Environ. Microbiol. *13*, 1956–1972.

Snyder, J.C., Bateson, M.M., Lavin, M., and Young, M.J. (2010). Use of cellular CRISPR (clusters of regularly interspaced short palindromic repeats) spacer-based microarrays for detection of viruses in environmental samples. Appl. Environ. Microbiol. *76*, 7251–7258.

Sorek, R., Kunin, V., and Hugenholtz, P. (2008). CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat. Rev. Microbiol. *6*, 181–186.

Suttle, C.A. (2007). Marine viruses—major players in the global ecosystem. Nat. Rev. Microbiol. *5*, 801–812.

Thingstad, T.F. (2000). Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. Limnol. Oceanogr. *45*, 1320–1328.

Thingstad, T., and Lignell, R. (1997). Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. Aquat. Microb. Ecol. *13*, 19–27.

Thingstad, T.F., Våge, S., Storesund, J.E., Sandaa, R.-A., and Giske, J. (2014). A theoretical analysis of how strain-specific viruses can control microbial species diversity. Proc. Natl. Acad. Sci. *111*, 7813–7818.

Tijdens, M., Hoogveld, H.L., Kamst-van Agterveld, M.P., Simis, S.G., Baudoux, A.-C., Laanbroek, H.J., and Gons, H.J. (2008). Population dynamics and diversity of viruses, bacteria and phytoplankton in a shallow eutrophic lake. Microb. Ecol. *56*, 29–42.

Tomasz, M. (1995). Mitomycin C: small, fast and deadly (but very selective). Chem. Biol. *2*, 575–579.

Tyson, G.W., and Banfield, J.F. (2008). Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. Environ. Microbiol. *10*, 200–207.

Villar-Argaiz, M., Medina-Sánchez, J.M., Cruz-Pizarro, L., and Carrillo, P. (2001). Inter-and intra-annual variability in the phytoplankton community of a high mountain lake: the influence of external (atmospheric) and internal (recycled) sources of phosphorus. Freshw. Biol. *46*, 1017–1034.

Weinbauer, M.G. (2004). Ecology of prokaryotic viruses. FEMS Microbiol. Rev. *28*, 127–181.

Weinberger, A.D., Sun, C.L., Pluciński, M.M., Denef, V.J., Thomas, B.C., Horvath, P., Barrangou, R., Gilmore, M.S., Getz, W.M., and Banfield, J.F. (2012). Persisting viral sequences shape microbial CRISPR-based immunity. PLoS Comput. Biol. *8*, e1002475.

Williamson, S., Houchin, L., McDaniel, L., and Paul, J. (2002). Seasonal variation in lysogeny as depicted by prophage induction in Tampa Bay, Florida. Appl. Environ. Microbiol. *68*, 4307–4314.

Williamson, S.J., Cary, S.C., Williamson, K.E., Helton, R.R., Bench, S.R., Winget, D., and Wommack, K.E. (2008). Lysogenic virus–host interactions predominate at deep-sea diffuse-flow hydrothermal vents. ISME J. *2*, 1112–1121.

Winter, C., Bouvier, T., Weinbauer, M.G., and Thingstad, T.F. (2010). Trade-offs between competition and defense specialists among unicellular planktonic organisms: the "killing the winner" hypothesis revisited. Microbiol. Mol. Biol. Rev. *74*, 42–57.

Wommack, K.E., and Colwell, R.R. (2000). Virioplankton: viruses in aquatic ecosystems. Microbiol. Mol. Biol. Rev. *64*, 69–114.

Wylie, K.M., Weinstock, G.M., and Storch, G.A. (2012). Emerging view of the human virome. Transl. Res. *160*, 283–290.

Yager, P.L., Connelly, T.L., Mortazavi, B., Wommack, K.E., Bano, N., Bauer, J.E., Opsahl, S., and Hollibaugh, J.T. (2001). Dynamic bacterial and viral response to an algal bloom at subzero temperatures. Limnol. Oceanogr. *46*, 790–801.

Zhong, X., Berdjeb, L., and Jacquet, S. (2013). Temporal dynamics and structure of picocyanobacteria and cyanomyoviruses in two large and deep peri-alpine lakes. FEMS Microbiol. Ecol. *86*, 312–326.

Chapter 2. Cyanobacterial and cyanophage populations show differing patterns between two biogeochemically similar neighboring lakes

2.1 Abstract

Viruses are diverse, dynamic and biogeochemically important components of aquatic ecosystems. Understanding what drives the dynamics between phage and host populations is critical to understanding the temporal cycling of phage infection *in situ*. Investigations in marine environments have discovered relationships between viral and host populations that appear predictable over time and in response to recurring annual conditions. Studies in freshwater environments have rarely observed such robust patterns. This study examines the dynamics of cyanobacteria and three lake-specific cyanophage populations in two neighboring morphologically and biogeochemically similar lakes. Based on comparisons of the abundance of members of each population in both lakes, cyanobacterial and cyanophage populations were only weakly correlated (Mantel test, $r=0.2182$, $p=0.002$). When populations were considered separately, the phage and host populations were correlated in only one of the lakes. Cyanophage populations exhibit similar dynamics between lakes, but not between two consecutive years; whereas cyanobacterial population dynamics show little consistency between lakes or years. Temperature significantly explained the variability in cyanophage and cyanobacterial population structure (CCA, $p<0.001$) suggesting that both populations follow seasonal patterns. Aside from temperature, differing environmental variables were shown to contribute to overall variability in population structure for the phage and cyanobacterial populations. Although the two lakes included in this study are highly similar, the dynamics of cyanobacterial and cyanophage populations between the two lakes appeared to be uncoupled.

2.2 Introduction

Viruses are ecologically important components of aquatic ecosystems with up to 10^{11} cells/ml in freshwater and 10^{8} cells/ml in marine environments. Most double stranded DNA viruses are known to be viruses of bacteria (Weinbauer, 2004). They are highly abundant, diverse and affect ecosystem processes through causing host mortality and through influencing community diversity and host metabolic processes during infection (Dinsdale et al., 2008; Suttle, 2007).

Picocyanobacteria are important components of aquatic habitats; *Prochlorococcus* and *Synechococcus* perform the majority of photosynthesis in the global ocean (Scanlan et al., 2009), and picocyanobacteria are prevalent in freshwater oligotrophic and mesotrophic systems (Callieri, 2008; Callieri and Stockner, 2002). Investigations into viruses of cyanobacteria (cyanophages) have helped uncover many interactions between viruses, hosts and the surrounding environment (Clokie and Mann, 2006; Lindell et al., 2007). For example, through the expression of analogs of host genes during infection, (e.g., photosystem genes *psbA* and *psbD*), cyanophages influence host processes and directly influence the cycling of carbon in aquatic habitats (Clokie and Mann, 2006; Lindell et al., 2007; Thompson et al., 2011). Previous studies have shown that 1 to 14% of *Synechococcus* are infected with phage in the ocean (Suttle and Chan, 1994), thus phage-mediated metabolic processes during infection and host mortality contribute significantly to carbon cycling in the microbial loop (Thompson et al., 2011).

As *Synechococcus* are found in diverse habitats (Callieri, 2008; Fahnenstiel and Carrick, 1992; Lu et al., 2001; Olson et al., 1990), it is not surprising that similar cyanophage genes are found in many different aquatic environments (Sullivan et al., 2008; Zhong and Jacquet, 2013). Through an analysis of a conserved biomarker gene, one study reported that genetically similar cyanobacterial myoviruses were found in diverse locations (Sullivan et al., 2008). Their prevalence in part is likely because of their ability to infect a large range of host cyanobacteria; cyanobacterial myoviruses are capable of infecting across species (Sullivan et al., 2003), and in one case, viruses

from a freshwater lake were shown to be capable of infecting a marine *Synechococcus* (Wilhelm et al., 2006). Understanding what drives the prevalence and dynamics between these virus-host populations sheds light on the role of viruses across habitats.

Studies comparing the abundance and assemblage structure of cyanophage and cyanobacterial populations show relationships of varying strength between the two groups. A study examining the diversity and abundance of both cyanobacteria and cyanophage found a strong relationship between the two groups and showed that the cyanophage populations were likely controlling cyanobacterial dynamics in the Red Sea (Mühling et al., 2005). A study in Norwegian coastal waters observed co-variation between cyanobacterial and cyanophage abundance, but no consistent predator-prey relationship (Sandaa and Larsen, 2006). Studies in freshwater environments that examine primary producer populations have been unable to find a strong connection between picocyanobacterial and viral assemblages. A study in a shallow eutrophic lake with filamentous cyanobacterial primary producers found bacterial abundances to be the only factor that correlated with the temporal variation in viral assemblage structure (Tijdens et al., 2008). A recent study investigating the assemblage structure of cyanobacterial and cyanophage populations within two French peri-alpine lakes showed no consistent relationship between cyanobacterial and cyanophage abundance or population structure in either lake examined (Zhong et al., 2013).

Similarly, investigations into the relationship between the overall viral and bacterial community structure revealed strong patterns in marine systems and variable patterns in freshwater systems. One study in the Gulf of Mexico found a relationship between the viral and bacterioplankton assemblage structure (Hewson et al., 2006). A comprehensive analysis examining bacterial, eukaryotic and T4-like myovirus assemblage structures off the coast of Southern California found a strong link between viruses and bacteria and revealed relationships between individuals from each group using a network analysis (Chow et al., 2013). A study in three

freshwater lakes using an approach similar to that of Hewson (2006) found a relationship between viral and bacterial assemblage structures in only one of the three lakes examined, and no relationship was noted when data from all lakes were considered together (Lymer et al., 2008). Several additional studies examine different aspects of the relationship between viruses and hosts in lakes of varying trophic states (Hardbower et al., 2012; Madan et al., 2005; Personnic et al., 2009; Tijdens et al., 2008). Aside from the well-characterized relationship between viral and bacterial abundance (Clasen et al., 2008), few universal patterns emerge across studies; however, shared patterns occur when systems of similar trophic states (enrichment level) are considered. For example, two studies examining the relationship between viral and bacterial populations within shallow eutrophic lakes observed correlations between viral and bacterial community structures (Hardbower et al., 2012; Tijdens et al., 2008). Similarly, two studies within meso- to oligotrophic lakes observed a high virus to bacteria ratio (VBR>80), four-fold higher than what is generally observed in aquatic environments (Madan et al., 2005; Personnic et al., 2009). However, the identical two studies had differing results about which factors contributed to the variability in viral abundance. One study found chlorophyll a measurements to significantly correlate (Madan et al., 2005), and the other found co-variation between virus like particles and heterotrophic bacteria (Personnic et al., 2009).

Therefore, the dynamics of viruses and bacteria in lakes are more variable and less predictable than in marine systems, and the lack of consistent trends is likely because of different physical and chemical environments between lakes. Indeed, the abundance and composition of picocyanobacteria within lakes appears to be dependent upon many variables such as the basin size, trophic status and water turbidity (Callieri, 2008). Similarly, the viral composition is observed to co-vary with temperature, the nitrate concentration and the phosphate concentration in freshwater lakes (Lymer and Lindström, 2010), parameters that can vary significantly between

lakes. Studies that compare lakes of differing trophic status not surprisingly display lake-dependent rather than universal relationships.

The current study seeks to reduce this variability through a comparison of the population dynamics of cyanobacteria and cyanophage in two biogeochemically similar, neighboring oligotrophic lakes. In addition to their physical similarities, the mixolimnion of the lakes are dominated by picocyanobacteria that mediate important biogeochemical processes and primary production. This population appears to be of low diversity, making it tractable through the use of molecular fingerprinting approaches. This study compares the dynamics of the cyanobacterial community to the abundance of three lake-specific cyanophage genotypes to address the hypothesis that in highly similar lakes, the dynamics of cyanobacteria and cyanophage will follow similar patterns.

Setting: Fayetteville Green Lake (GL) and Round Lake (RL) are located in Green Lakes State Park in Fayetteville, NY. Both lakes are meromictic (permanently stratified) with an oligotrophic, oxic mixolimnion and an anoxic, sulfurous (gypsum rich) monimolimnion. They are approximately 500 m in diameter and 50 m deep. Historical studies report high concentrations of ions such as calcium (average 10.48 mM), magnesium (average 2.96 mM), sulfate (average 11.66 mM) and bicarbonate (average 3.28 mM) (Brunskill and Ludlam, 1969). These two lakes are isolated within a single small catchment basin less than 5 km^2 in area (Hilfinger IV et al., 2001). A small stream connecting the two lakes flows from RL to GL. Mixolimnion residence times are estimated to be <2 years (Brunskill and Ludlam, 1969; Takahashi et al., 1968; Torgersen et al., 1981). The dominant primary producers are picocyanobacteria of the genus *Synechococcus* which are responsible for calcite precipitation in the lake. This precipitation has led to the formation of carbonate bioherms around the perimeter of GL and cause whiting events every spring in which calcium carbonate is precipitated out of the water column (Schultze-Lam et al., 1992; Thompson and Ferris, 1990; Thompson et al., 1997). These features are not shared with other near-by lakes, making these lakes biogeochemically distinct. A previous study of picocyanobacterial abundance within GL showed a seasonal dynamic in the surface waters: an increase in population density from May to June, followed by a mid-summer crash and a return to abundance by late August/early September (Thompson et al., 1997). A reported viral metagenomic sequence library from FGL showed that cyanobacterial viruses (cyanophage), genetically similar to phages known to infect *Synechococcus* and *Prochlorococcus*,,comprise roughly 30% of identifiable viral sequences (Hewson et al., 2012), suggesting that cyanophage likely play an important role in GL.

2.3 Methods

Time Series Sampling: Lakes were sampled weekly from April to early August in 2012 (4/27 to 8/9), once in the beginning of September (9/7) and seven times throughout the summer of 2013 from April to August (4/11, 4/30, 5/22, 6/17, 7/9, 8/5, 8/31). The initial time point for 2012 was selected based on historical data showing a *Synechococcus* bloom in the surface waters in mid-May (Thompson et al., 1997) and sampled weekly for the summer season to capture the return to abundance in late August/early September. In 2013, sample initiation began as close to ice out as logistically possible and spanned a comparable period of time to the 2012 sampling season. Green Lake was sampled from off the shelf of the area called "Dead Man's Point" on the eastern side of the lake. Round Lake was sampled on the southern-most side of the lake. Water was collected 2 m from the GL/RL shelf into the surface waters using a custom pump and tubing setup. Samples were withdrawn from ~10 cm below the surface. Acid-rinsed bottles were triple rinsed with sample water and then filled with lake water and stored in the dark at lake temperature for 2 to 6 hours before processing. Samples were collected for bacterial (500-1000 ml, 10 um polycarbonate (Isopore; Millipore) pre-filter collected onto a 0.2 um polyvinyldifluoride (Durapore; Millipore) filter) and viral (50-200 ml, 0.2 um PVDF pre-filter collected on a 0.02 um anotop (Whatman) filter) DNA using a peristaltic pump and swinnex filter holders for the 10 um and 0.2 um filters. In total, 50 ml of 0.2 um filtered water from each sample was stored at -80°C for nutrient analysis.

Measurements: Nitrate/Nitrite was measured at University of California Santa Barbara via Flow Injection Analysis at the UCSB nutrient analysis facility (William J. Clinton, http://msi.ucsb.edu/services/analytical-lab/instruments/flow-injection-analyzer). Soluble reactive phosphate (SRP) was measured on an autoanalyzer in the Cornell Soil and Water Lab via colorometric analysis (U. S. Environmental Protection Agency, 1983). Temperature, conductivity and dissolved oxygen were measured immediately before sample collection using a YSI probe (YSI

Pro 2030). pH was measured at Cornell using a pH 510 series electrode (Oakton Instruments). Virus like particles (VLP) and bacterial abundances were determined by SYBR Green 1 staining and epifluorescence microscopy as previously described (Noble, 2001; Patel et al., 2007). Picocyanobacterial abundance was determined via filtration of 15 to 40 ml of Formalin-fixed sample over a 0.2 um black isopore filter and visualized via epifluorescence microscopy using a CY3 filter cube (513-556 excitation, 570-613 emmision) at 400× magnification (Waterbury et al., 1979).

Precipitation and air temperature data were taken from NOAA's Climate Data Center website (NOAA). Data presented in this study were collected at the State University of New York's Environmental Science and Forestry (SUNY ESF) weather station, approximately 10 miles away from Green Lakes State Park.

DNA Extraction: Bacterial DNA was extracted from filters using Zymo Fungal/Bacterial DNA extraction kit (Zymo Research, Irvine, CA). Filters were placed in sterile bead beating tubes with lysis buffer and agitated using a vortexer with a bead beating attachment at maximum speed for 10 minutes. Subsequent steps followed the protocol per the manufacturer's instructions.

Viral DNA was extracted from anotop filters using a Zymo Viral DNA extraction kit with a modified protocol. Anotop filters were thawed and 800 µl of ZR viral buffer was introduced into the filter using a sterile 3 ml syringe. When the buffer filled the filter, anotops were plugged with a sterile flame-sealed micropipette tip, vortexed for 5 seconds and then incubated at room temperature for 10 minutes. After 10 minutes, the majority of the viral buffer was removed from the filter back into the syringe using backpressure, and the liquid in the syringe was transferred to a sterile tube. The filter was then pierced with a sterile sealed micropipette tip and the remaining liquid was pushed through the filter and into the same sterile tube. The extraction then proceeded with the expunged viral buffer according to the manufacturer's protocol.

Design and implementation of qPCR primers: A clone library of g20 sequences from the surface waters of GL and RL was constructed to identify which populations of cyanobacterial myoviruses to target for qPCR. Polymerase chain reaction (PCR) was used to amplify a conserved cyanobacterial myovirus portal protein gene (g20) from viral DNA collected from RL and GL during the summer of 2012 and GL in 2013 using cyanomyovirus specific primers cps1.1 (5′-GTAGWATWTTYTAYATTGAYGTWGG-3′) and cps8.1 (5′-ARTAYTTDCCDAYRWAWGGWTC-3′) (Sullivan et al., 2008). PCR reactions contained 1× reaction buffer, 3.5 mM $MgCl_2$, 0.05 U/µl Taq Polymerase (Promega), 1 mM dNTPs (Promega PCR Nucleotide Mix), 2 µl extracted viral DNA, 20 pg/µl bovine serum albumin, and 0.05 µM of each primer. Reactions began at 94°C for 3 minutes and then cycled through 35 rounds of 94°C melting temperature for 1 minute, 35°C annealing temperature for 1 minute, 73°C extension for 1 minute. After 35 cycles, the reactions sat at 73°C for 4 minutes for a final extension. Sequenced time points were selected based on the success of amplification of a PCR product of the expected size. Amplified DNA was gel purified, cloned and sequenced, resulting in 109 clones from Round Lake and Green Lake from 2012, and 42 clones from Green Lake in 2013.

DNA sequences were trimmed and curated using CLC Bio's Main Workbench. g20 sequences from the RL and GL water column were aligned in MEGA (Kumar et al., 2008) using MUSCLE (codon-based alignment, default parameters, Edgar, 2004) and clustered at 98% nucleotide identity using mothur (Schloss et al., 2009). Taqman qPCR primers were designed with the aid of Primer3 (Untergasser et al., 2007) around the consensus sequences of the two most highly represented clusters (primer sets RLCP2A and RLCP1) and a third OTU (RLCP4) selected based on its distance from the other two selected OTUs. These three primer sets were designed with the goal of capturing both highly abundant and lower abundance populations. Primer sequences were compared to NCBI's nr database and to other cluster representatives within the g20 library to verify primer specificity. Primers include a forward and reverse sequence and a

TaqMan probe which ensures specificity over the 20 to 25 bp sequence in the middle of the PCR amplicon. Triplicate qPCR reactions for each primer set were conducted for each viral DNA sample, with 5 standards per run ranging from 10^8 to 10^1 copies per reaction. Reaction mixes contained 1× BioRad iTaq master mix, 2 µM forward and reverse primer, 0.4 µM probe and 1 µl extracted viral DNA. qPCR reactions were run on a ABI 7800 thermocycler. Reactions began with 2 minutes at 50°C, 2 minutes at 95°C, then cycled through 15 seconds at 95°C, and 1 min at the Tm for each primer set (Table 2.1) 40 times. The abundance of each OTU was calculated as copies per milliliter of lake water.

Table 2.1. TaqMan qPCR primers designed for this study

Primer Set	Tm	Forward	Reverse	Probe
RLCP1	63	ACAGGTGAGTTGCGTGATGA	TGTGCCTTTACCACCCTCAC	TGCTAGAAGACTTCTGGTTACCACGCC
RLCP2A	60	CGTGAAGGTGGAAAAGGTACA	ACCACCTTGTGGTTCTAAACG	CCAGCTGGTCAAAATCTTGGCCA
RLCP4	58	CAAAATCTTGGAGAACTTGCTG	ATGAACGRCCAAGATTGAAA	CGAATYGATGTTGGTGGTGGTGG

Amplification for Cyanobacterial ARISA: The picocyanobacterial population was assessed using cyanobacteria-specific automated ribosomal intergenic spacer analysis (cARISA). PCR reactions contained 1× reaction buffer, 1.5 mM $MgCl_2$, 0.05 U/µl Taq Polymerase (Promega), 1 mM dNTPs (Promega PCR Nucleotide Mix), 2 µl extracted microbial DNA, 20 pg/µl bovine serum albumin, and 0.05 uM of primers 1247F (5'-CGTACTACAATGGTTGGG-3'; (Rocap et al., 2002)) and 23S-125R (5'-GGGTT[C/G/T]CCCCATTC [A/G]G-3'; (Fisher and Triplett, 1999)) labeled with a 5'-TET (6-carboxy-1,4-dichloro-2',7'-dichlorofluorescein, a fluorescent probe used to detect DNA fragments). Reactions began at 4°C for 4 minutes and then cycled through 30 rounds of 94°C melting temperature for 1min, 53.5°C annealing temperature for 1 minute, 72°C extension for 3 minutes. After 30 cycles, the reactions sat at 72°C for 10 minutes for a final extension. All dominant bands within cARISA PCR reactions were cloned and sequenced to identify the associated DNA sequence. PCR bands were gel purified, cloned using the P-GEM T-Easy cloning vector system (Promega) and sequenced at the Cornell Biotechnology Resource Center.

Phylogenetic Analysis of Representative ITS Sequences: Representative GL ITS sequences were combined with closely related sequences identified through a BLASTn against NCBI's RefSeq database (Altschul et al., 1997). Additional sequences from the PopSet from Rocap (2002) were added as marine representatives and are also available through NCBI (See Appendix 1 Table A1.1). Sequences were combined and aligned in MEGA (Kumar et al., 2008; Tamura et al., 2013) using MUSCLE (Gap Open Penalty of -300, UPGMB Clustering Method, Minimum Diagonal Length of 24, (Edgar, 2004)) and hand curated to remove gaps. The analysis involved 52 nucleotide sequences. A total of 518 positions were in the final dataset. The phylogenetic tree was constructed in MEGA using the Maximum Likelihood model with 100 iterations for bootstrap support (Tamura et al., 2013).

Automated analysis of cARISA reactions: For automated fragment size analyses, duplicate PCR reactions for each sample were pooled and cleaned using the Zymo Research clean and concentrator kit (Zymo Research, Santa Barbara, CA), combined with a ROX-labeled internal size standard ranging from 200 to 1350 bp (Bioventures, Inc) and submitted to the Cornell Biotechnology Resource Center for DNA fragment analysis on an ABI 377XL capillary gel electrophoretic instrument. Electropherograms were analyzed using Peakscanner software (ABI). Electropherograms were manually curated and picocyanobacterial peaks were identified based on the predicted length of sequenced ITS amplicons. Peak size distributions were selected based upon manual observations of peak abundance compared to the results from the sequenced reactions. Size bins used in subsequent analyses are reviewed in Table 2.2. A relative representation of each picocyanobacterial OTU was determined by comparing the total area of each cyanobacterial peak between all runs pertaining to the time point and location to the total area of picocyanobacterial peaks within all runs from the identical time point. Relative representation of each cyanobacterial peak was multiplied by picocyanobacterial abundance to determine a semi-quantitative

representation of each OTU at each time point and location. Data processing, analyses and visualization was conducted in R (Team, 2012).

Statistical analyses of time series data: Cyanobacterial type abundance (cells/ml) and cyanophage abundance (copies/ml) were log(x+1) transformed to obtain a normal distribution for multivariate analyses. Bray-Curtis dissimilarity matrices were constructed from the log(x+1)-normalized semi-quantitative abundance of cyanobacterial types and the log(x+1)-normalized abundance of the three cyanophage types. To test the relationship between cyanobacterial and cyanophage populations, Bray-Curtis dissimilarity matrices were compared via Mantel tests using Pearson's correlation coefficient and 1000 permutations. Additional Mantel tests of lake-specific and year-specific populations were conducted in the same manner.

Canonical Correspondence Analysis (CCA) was employed to examine the relationships of cyanobacterial and cyanophage populations with physical and chemical conditions of GL and RL as constraining axes. Physical, chemical and counts data considered in these analyses (nitrate concentration, nitrite concentration, soluble reactive phosphorus, temperature, average rainfall, bacterial abundance, viral abundance and cyanobacterial abundance (for the cyanophage data only)) were centered around their means with standardized variances (z-score transformation). CCAs were subjected to forward selection until an optimal combination of constraining variables explained the maximum amount of variability within the population structure. Significance of influence of physical and chemical data on constrained axes was determined via permutation tests using the 'anova' function within the vegan package (Oksanen et al., 2007). In addition to comparison to physical, chemical and counts data, CCAs were conducted on cyanophage populations using chi square-transformed cyanobacterial abundance data as the constraining variables, and CCA on cyanobacterial populations using chi-square transformed cyanophage abundance data as the constraining variables. All analyses were conducted in R (Team, 2012) using the vegan (Oksanen et al., 2007) and MASS packages (Ripley, 2011).

2.4 Results

Lake Conditions

Sampling of Green Lake (GL) and Round Lake (RL) was performed in two consecutive years, over a period of 5 months per year. Samples were taken in duplicate from one location in each of GL and RL. Between the two years, the region experienced vastly different precipitation patterns with low levels of rainfall in the summer of 2012 and heavy rains during the summer of 2013 (Figure 2.1).

Nutrient concentrations varied over the course of the two seasons and between lakes (Figure 2.2). Nitrate and nitrite was twice as high in Green Lake compared to Round Lake, and concentrations remained relatively stable over the course of the summer months. SRP was similar between the two lakes and years. Dissolved oxygen remained between 8 and 12 mg/L for all samples except three time points, which fell below 6 mg/L. pH measurements fell between 7.6 and 8.0 in both lakes.

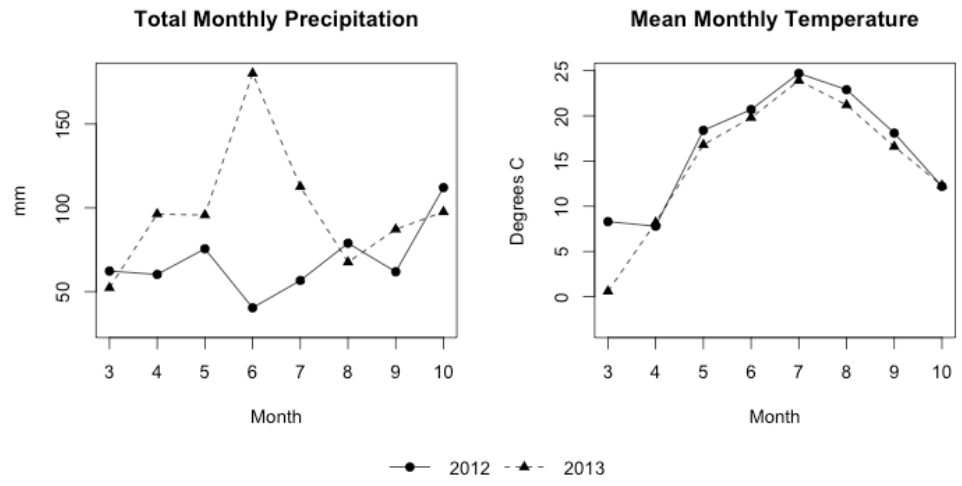


Figure 2.1 Total Monthly Precipitation (left) and mean monthly air temperature (right) for the Syracuse region from NOAA's National Climate Data Center (NOAA NCDC).

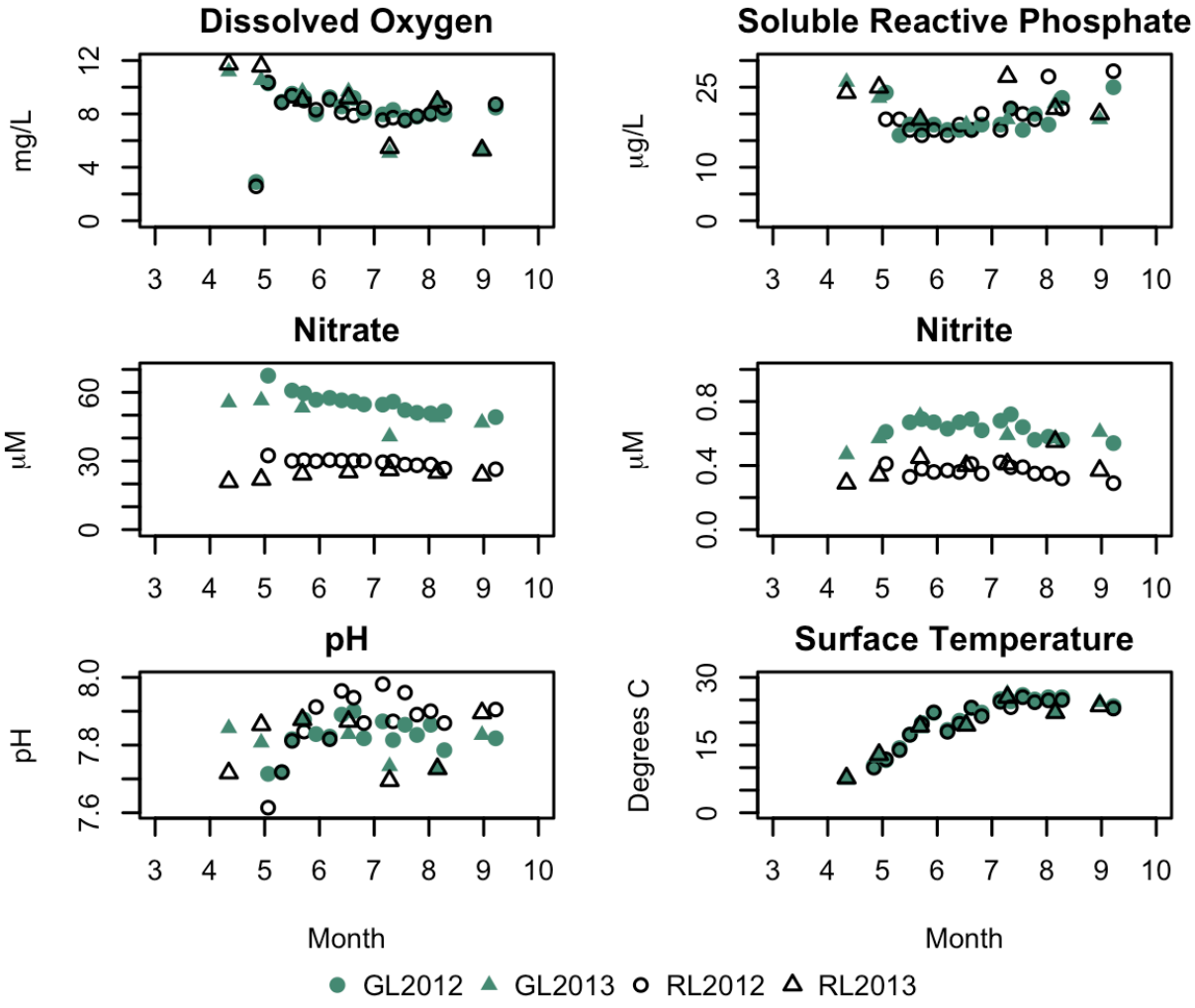


Figure 2.2 Physical and chemical measurements over the sampling period in GL and RL.

Bacterial and Virus-like particle abundance

Bacterial and virus like particle (VLP) abundance decreased throughout the summer of 2012 and 2013, with highest levels in May, decreasing to August (Figure 2.3). The VLP abundance was different in 2013 compared to 2012. In 2013, RL had consistently higher levels of VLP ($1.696*10^7$ +/- $5.597*10^6$ VLP/ml) compared to GL ($7.330*10^6$ +/- $5.178*10^6$ VLP/ml, paired t-test $p<0.001$, Figure 2.3). Regardless, a relationship was noted between bacterial and VLP abundance; the log-transformed abundance of VLP and bacteria fit well into a linear model (slope= 3.5750,$p<0.05$, $R^2=0.48$) and confirm this relationship. The relationship between viruses and bacteria is consistent with previous studies in aquatic systems (Clasen et al., 2008).

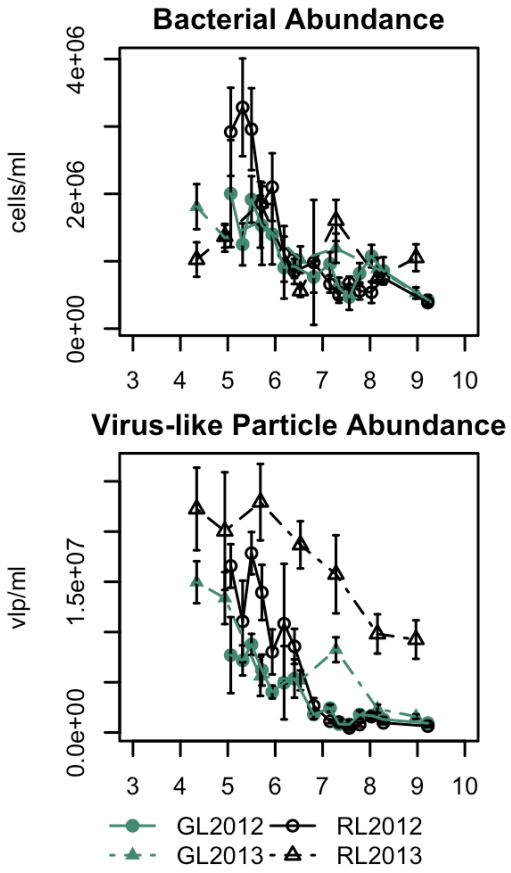


Figure 2.3 Bacterial (top) and virus like particle (VLP, bottom) abundance in GL and RL over the two sampling seasons

Picocyanobacterial abundance and community structure

The picocyanobacterial populations in RL and GL were characterized and monitored over time by direct cell counts of phycoerythrin-containing cells (via epifluorescence microscopy) for overall cyanobacterial abundance, and using cyanobacteria-specific Automated Ribosomal Intergenic Spacer Analysis (cARISA) to determine population structure. cARISA has been previously used to characterize bioherm-associated cyanobacterial populations in GL (Wilhelm and Hewson, 2012) and in coastal marine systems (Chamberlain et al., 2014) and using more general bacteria-specific primers to characterize bacterial communities in diverse habitats (Brown et al., 2005; Fisher and Triplett, 1999; Hewson et al., 2006; Ranjard et al., 2001).

Four populations of picocyanobacteria were observed in GL and RL, with intergenic spacer amplicons ranging from 1129 bp to 1360 bp (Table 2.2). All four types fell into the genus *Synechococcus* (Figure 2.4). Each group contained sequences that varied in size by 2 to 8 bp, but encompassed a shared peak that could not be distinguished at the large bp window of cARISA electropherograms. All sequences that fell within each group were greater than 99% identical in nucleotide sequence over the 1129 to 1360 bp amplicon. Fragment sizes were consistently smaller in measured length within cARISA electropherograms than their sequenced length in clone libraries (15 to 40 bp shorter in size within electropherograms compared to sequenced clones, see Table 2.2). This result is consistent withobservations of large amplicon electropherograms outputs considered in previous studies (Brown et al., 2005; Hewson and Fuhrman, 2006) . Populations are phylogenetically distinct, with similarity to environmental sequences from the Baltic Sea (Haverkamp et al., 2008), East Tibetan Lakes (Wu et al., 2010), European subalpine lakes (Crosbie et al., 2003; Ernst et al., 2003) and other freshwater locations (Figure 2.5). Closest representative genomes of cultivated organisms in RefSeq are distantly related (75-79% nucleotide identity) to sequences derived in this study (Figure 2.5, Table 2.2).

Of the four cyanobacterial phylotypes, types 2 and 3 were consistently present over the season. Type 1 cyanobacteria were present in the early season, but absent in mid-summer, and type 4 cyanobacteria exhibited low abundance throughout the summer in both lakes, over both years (Figure 2.4, left panels).

Enumerated cyanomyovirus populations

This study examined the abundance of three cyanomyovirus genotypes over the course of two seasons by looking at the copy number of three specific g20 OTUs using qPCR. Three different cyanomyovirus genotypes were selected for quantification: Cyanomyovirus genotypes RLCP1 and RLCP2a were the most highly represented genotypes within the constructed clone libraries; RLCP4 was selected as a third, more distantly related and less abundant within the clone libraries (Figure 2.6).

These three populations of cyanophage exhibited generally shared trends in abundance over the sampling season in both lakes. In 2012, all three types were present during the earlier months and disappeared by mid-summer. In 2013, peaks in abundance varied between populations, RLCP4 peaked early in the season and RLCP1 and RLCP2a were present throughout the summer (Figure 2.4, right panels).

Corresponding rates of change for cyanobacterial and cyanophage populations are plotted on Figure A1.2 in Appendix 1.

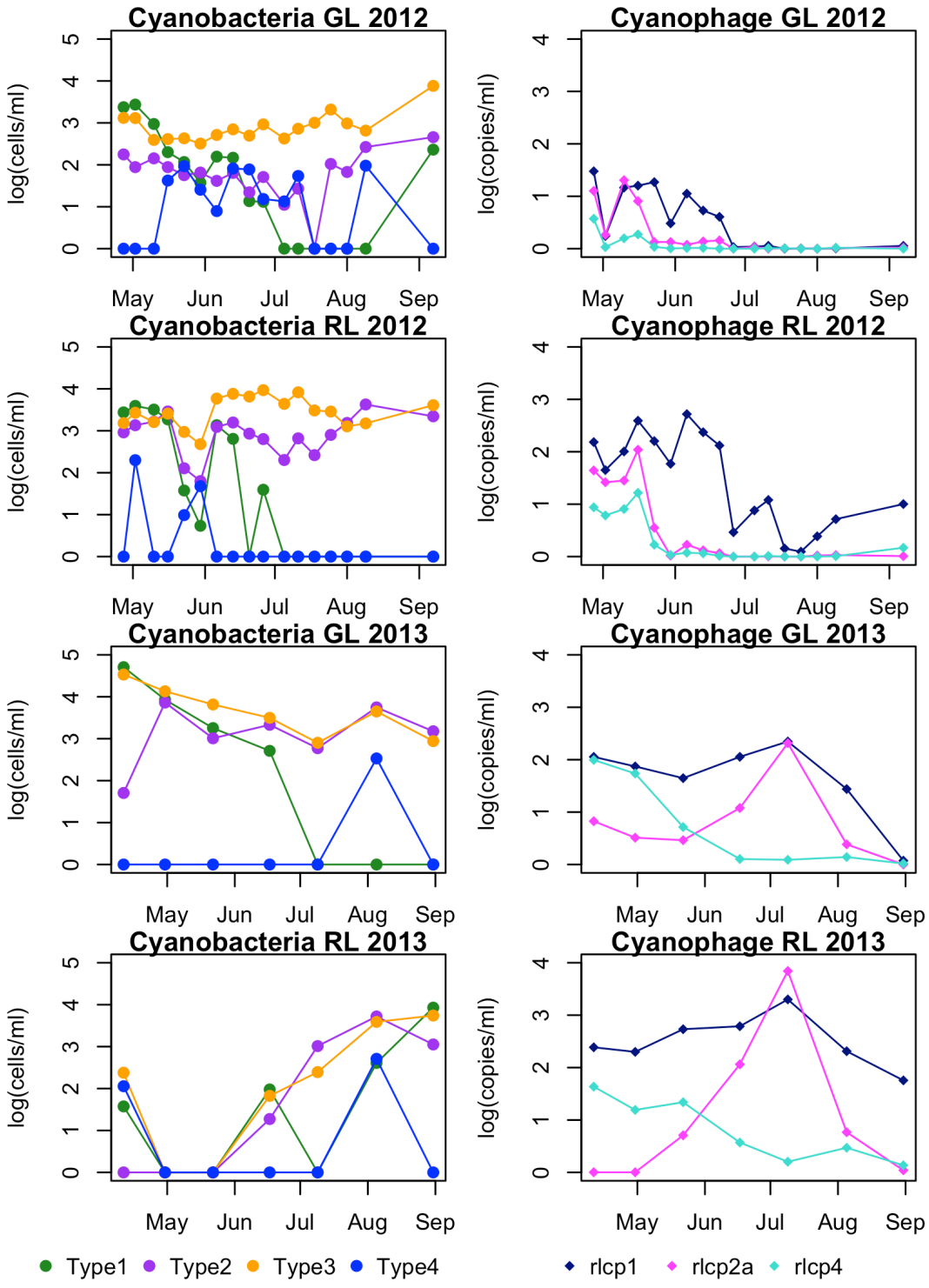


Figure 2.4 Log abundance of cyanobacterial populations determined by multiplying the relative abundance of each cyanobacterial type, determined using ARISA, by the overall abundance of phycoerythryn containing cells, determined via fluorescence microscopy cell counts (left panels) and abundance of three cyanophage populations determined using Taqman quantitative PCR (right panels) .

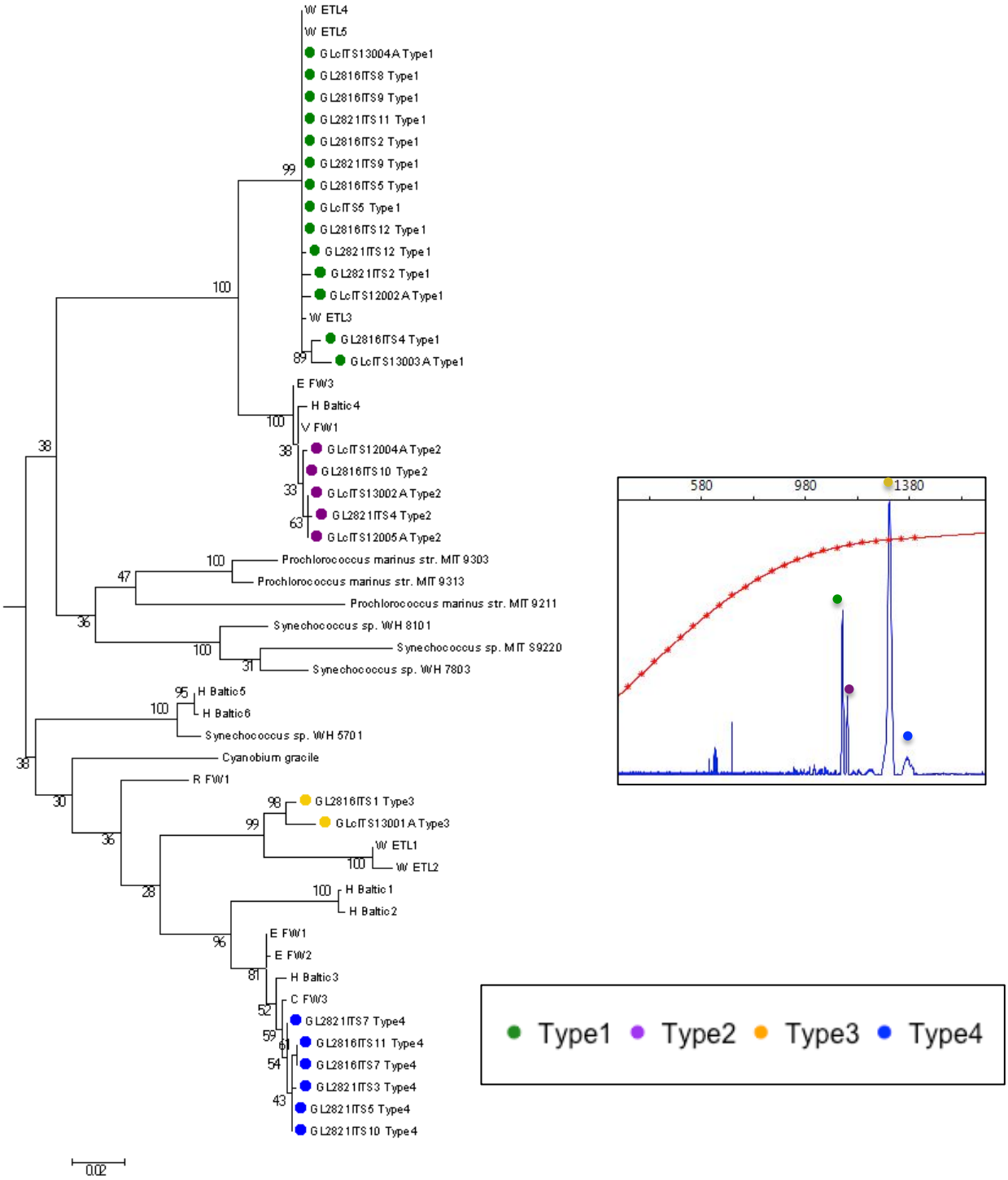


Figure 2.5 Maximum Likelihood tree of ITS nucleic acid sequences from Green Lake with closest matches available through NR and marine representatives (Appendix 1, Table A1.1). Cyanobacterial types examined in this paper are noted with a circle next to the name. The analysis involved 52 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 518 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura 2013). Inset electropherogram provides example of all four cyanobacterial types used for cARISA analysis with colored circles indicating the four phylotypes. The red line indicates a size standard, with red points denoting every 50 base pairs from 200bp to 1350 bp.

Table 2.2 Characterization of the four cARISA peaks present in GL and RL identified as belonging to picocyanobacteria via best BLAST match. Here, best BLAST matches of each cARISA peak is compared to the non-redundant database at NCBI (nr) and the NCBI's RefSeq

cARISA Band			nr best BLAST match				RefSeq Genomes best BLAST match		
Type	**Size Range (bp)**	**cARISA Size**	**GenBank Accession**	**Name/ Location**	**Nucleotide Identity/ Bit Score**	**Citation**	**GenBank Accession**	**Organism Name**	**Nucleotide Identity/ Bit Score**
1	1129-1137	1108-1125	FJ596205.1	Environmental Clone/ Tibetan Lakes	0.99/1415	Wu et al 2010	NC_009482.1	Synechococcus sp. RCC307	0.79/1016
2	1146-1148	1130-1142	FJ596227.1	Environmental Clone/ Tibetan Lakes	0.98/1807	Wu et al 2010	NC_009482.1	Synechococcus sp. RCC307	0.79/1052
3	1308-1310	1240-1304	FJ596210.1	Environmental Clone/ Tibetan Lakes	0.93/1378	Wu et al 2010	NC_008820.1	Prochlorococcus marinus str. MIT 9303	0.75/922
4	1355-1361	1310-1336	AM709629.1	Synechococcus rubescens SAG 3.81	0.98/2364	Marin et al. 2007	NC_019675.1	Cyanobium gracile PCC 6307	0.75/955

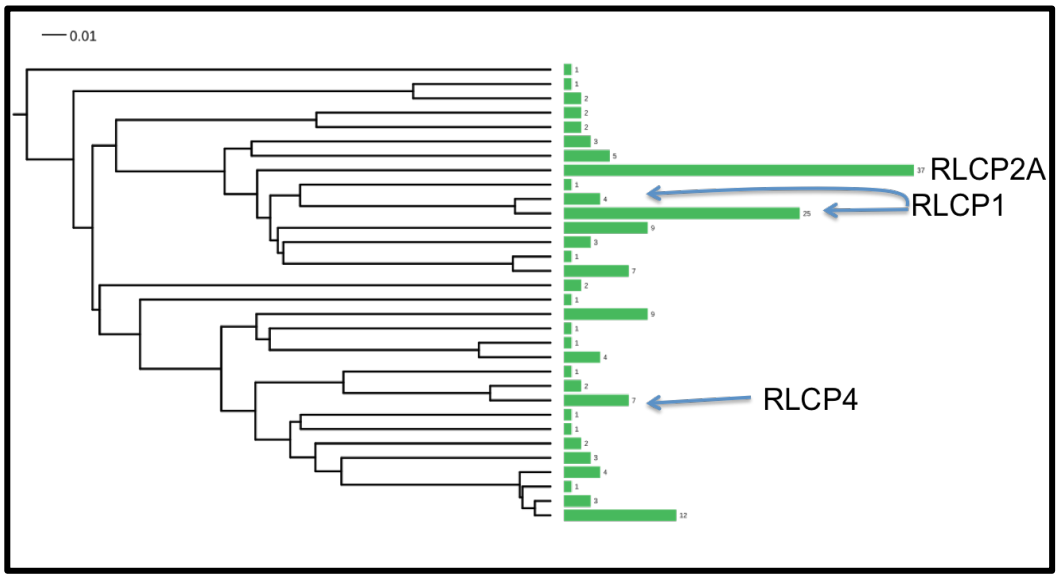


Figure 2.6. Dendrogram of g20 nucleic acid sequences from the GL and RL water column clustered at 98% nucleic acid identity using the clustering algorithm available in mothur (Schloss 2009), visualized in iTOL (Letunic 2007). The scale bar represents the difference in nucleotide identity. Green bars represent the number of sequences within each cluster. Clusters from which qPCR primers were designed are noted.

Relationships between cyanophage and cyanobacterial populations within and between lakes

Mantel tests probing the relationship between viral and bacterial community composition revealed lake-dependent relationships. Overall, when comparing Bray-Curtis dissimilarity matrices of cyanobacterial and cyanophage populations at all timepoints in both lakes, the populations were significantly but weakly correlated (Mantel test, r=0.2182, p=0.002). This was also true when comparing only populations sampled in 2012, and only populations from GL were considered separately. The strongest correlation between cyanobacterial and cyanophage populations was noted for GL 2012 (Table 2.3; Mantel test; r=0.4351, p=0.001). When the populations from Round Lake were considered alone, no significant relationship was noted (Table 2.3). A significant correlation was noted between cyanophage populations in GL and RL over time, but no similar relationship for cyanobacteria was noted between the two lakes (Table 2.3; Mantel test; r=0.7711, p=0.001).

To investigate the possibility of observable predator-prey interactions between cyanobacterial and cyanophage populations, the rate of change of each population per day was calculated for each population in each lake (Appendix 1, Figure A1.2). In addition, the abundance of cyanobacteria was plotted against the combined total abundance of the three cyanophage populations. No consistent patterns were observed using these parameters, further showing that the cyanobacterial and cyanophage populations are uncoupled (Appendix 1, Figure A1.3).

Table 2.3 Mantel test comparing examining Bray-Curtis dissimilarity matrices from various population subsets, expressed as a Pearson correlation coefficient with significance in parentheses . Significant correlations are shown in bold.

Comparison	Mantel Statistic (Significance)
all Cyanobacteria -- all Cyanophage	**0.2182(0.002)**
GL Cyanobacteria -- GL Cyanophage	**0.3362(0.002)**
RL Cyanobacteria -- RL Cyanophage	0.1078(0.166)
2012 Cyanobacteria -- 2012 Cyanophage	**0.2483(0.002)**
2013 Cyanobacteria -- 2013 Cyanophage	0.08085(0.278)
GL 2012 Cyanobacteria -- GL 2012 Cyanophage	**0.4351(0.001)**
RL 2012 Cyanobacteria -- RL 2012 Cyanophage	0.2007(0.055)
GL Cyanobacteria -- RL Cyanobacteria	0.1786(0.098)
GL Cyanophage -- RL Cyanophage	**0.7711(0.001)**

Physical and Chemical Influences on Population Structure

The influences of physical and chemical variables on population structure were investigated using forward selecting CCA. For cyanobacteria, dissolved oxygen, temperature, pH and soluble reactive phosphorus contributed to 46.38% of population variability (Figure 2.7A, Table 2.4). Dissolved oxygen, pH, temperature, total monthly precipitation (TPCP), bacterial abundance (bac) and nitrate concentration contributed to 86.51% of the variability in cyanophage population structure, with nitrate being the least significant contributor (Figure 2.7B, Table 2.4).

Influences of host and phage populations on corresponding population structure

The influence of cyanophage population structure on cyanobacterial dynamics and vice versa was tested through CCA of each population, using the other as constraining variables. Based on the cyanophage plot, cyanobacterial constraining axes accounted for 37% of the variability in the cyanophage population structure. Two out of the four cyanobacterial types contributed significantly (Table 2.4). Little contribution was noted from cyanophage populations to the variability in cyanobacterial population structure, with no cyanophage types significantly contributing to cyanobacterial variability (Table 2.4).

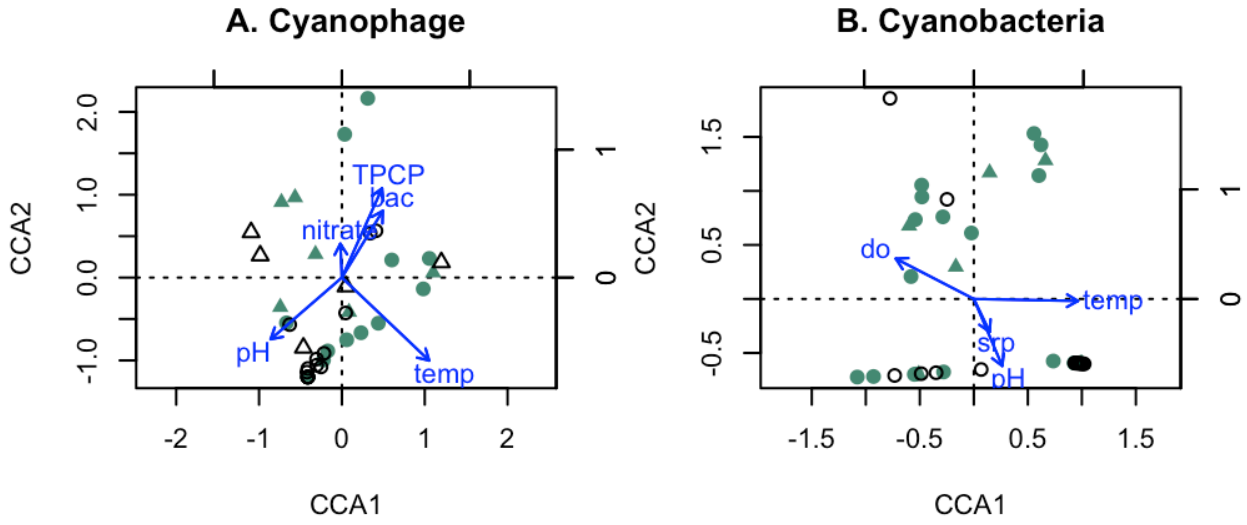


Figure 2.7. Cannonical Correspondence Analyses (CCA) of cyanobacterial populations determined via cell counts and ARISA data, and cyanophage populations determined using qPCR using forward selection of physical and chemical measurements as constraining variables. Constraining variables selected for the cyanophage model include total monthly precipitation (TPCP), bacterial abundance (bac), temperature (temp), nitrate and pH (A). Constraining variables selected for the cyanobacterial model include dissolved oxygen (do), soluble reactive phosphate (srp), pH and temperature (B).

Table 2.4 Statistics associated with CCA plots in Figure 2.7, and CCAs comparing cyanobacterial and cyanophage populations. Contributing variables for the cyanophage CCA were temperature (temp), pH, average monthly precipitation (TPCP), bacterial abundance (bac) and nitrate. Contributing variables for the cyanobacterial CCA included temperature, pH, dissolved oxygen (do) and soluble reactive phosphate (srp).

Analysis	Proportion Contribution of Constrained Axes	CCA1 Eigenvalue	CCA2 Eigenvalue	Contributing Variables
Cyanophage CCA, Forward Selected Constraints	0.8498	0.1679	0.133	temp**, pH**, TPCP**, bac**, nitrate~
Cyanobacteria CCA, Forward Selected Constraints	0.4638	0.09917	0.05015	temp**, pH*, do*, srp*
Cyanophage CCA, Cyanobacterial Constraints	0.3709	0.105	0.002	Cyano1, Cyano2**, Cyano3*, Cyano4
Cyanobacterial CCA, Cyanophage Constraints	0.0862	0.0024	0.00041	RLCP1, RLCP2A, RLCP4
**** <0.01, *<0.05, ~<1**				

2.5 Discussion

This study examines the dynamics of cyanobacteria within GL and RL through a semi-quantitative measurement of four distinct cyanobacterial populations throughout two summer sampling seasons. In addition, the study follows three lake-specific cyanophage populations over the identical period. Because RL and GL neighbor each other and are morphologically and biogeochemically similar, and because they are physically connected, their cyanobacterial and cyanophage populations were expected to exhibit shared population dynamics throughout the sampling period. This was not the case. Differences between the lakes and years reveals insights into what may contribute to the differing dynamics.

Several physical, chemical and biological characteristics differed between GL and RL. GL has nearly double the concentration of nitrate compared to RL (Figure 2.2). The reason for this is not known. A historical study of GL and RL found similar concentrations of nitrate in the GL and RL surface waters (Brunskill, 1968).

Because of permanent stratification and low levels of nutrients in the surface waters, storm water runoff may be an important contributor to lake nutrients in the mixolimnion. This is another factor that may vary between lakes as both lakes receive run-off from gullies in their respective basins, however GL receives stream inputs from RL whereas RL receives inputs from a transient stream at the eastern side of Round Lake (opposite the stream that flows from RL to GL).

A log-linear relationship was noted between virus like particle and bacterial abundance that is shared between the lakes. The virus to bacteria ratio (VBR), a common metric used to assess the two populations in aquatic environments, was low in both GL and RL, ranging from less than 1 (0.67) to 33, with a mean of 5.9 (Appendix 1, Figure A1.1). The abundance of virus like particles was significantly higher in RL 2013, which was reflected in an increased VBR that peaked in mid-summer (Figure A1.1). Low ratios of viruses to bacteria have also been observed in deep-sea sediments (Danovaro and Serresi, 2000; Danovaro et al., 2008), but in these cases, a high

abundance of both viruses and bacteria is noted. Other instances of observed low abundances of viruses relative to bacteria include oxygen minimum zones (Cassman et al., 2012) and extreme aquatic systems (Siering et al., 2013). The deep ocean possesses low concentrations of viruses relative to other aquatic habitats, but bacterial abundance is also low and the ratio of viruses to bacteria remains >10 (Parada et al., 2007). In GL and RL, low ratios of viruses to bacteria coincided with a decrease in both populations during the time series and the peak in the VBR in RL coincides with a month of heavy precipitation.

Direct comparison of cyanobacterial and cyanophage populations

Cyanobacterial and cyanophage populations are correlated when considering all samples (Mantel test, $r=0.22$, $p<0.002$, Table 2.3). However, the relationship between cyanobacteria and cyanophage appear different between GL and RL. A significant relationship was noted between cyanobacteria and cyanophage when GL populations are considered on their own, but RL populations do not show this robust correlation (Mantel tests, Table 2.3). These results are similar to results from a study of the bacterial and phage populations within three peri-alpine lakes. This study found a relationship between the structure of bacterial and viral populations in only one of the three lakes examined (Lymer et al., 2008). Of the three lakes examined by Lymer (2008), two were geochemically similar and two were geographically similar. In the current study, GL and RL are morphologically, biogeochemically, and geographically similar.

The current study observes a relationship between cyanobacterial and cyanophage populations that is strongest during the summer of 2012 in Green Lake. Other periods of sampling exhibited no apparent relationship between the abundance of the measured cyanobacterial and cyanophage populations. In addition, there appeared to be no clear predator-prey relationships between these groups (Figure A1.3), suggesting that forces beyond interactions between cyanobacteria and cyanophage are driving the dynamics of these two populations. One possibility for the lack of predator prey interactions is that the abundance of the host population is simply too

low to be influenced by predation. This is possible because lytic phage infection is contact-rate dependent. With few available hosts and low concentrations of viruses there is less probability that a phage and host will contact each other, leading to an inability of the phage population to expand at a high enough rate to influence host dynamics. Previous studies have observed such a pattern, in which the initial abundance of the viral population determined whether or not the viral population would expand during a mesocosm incubation (Wilcox and Fuhrman, 1994). This study found that if viruses were below a certain threshold at experiment onset, no viral production was observed regardless of bacterial growth. The contact rate of viruses and bacteria in an environment is dependent upon the concentrations of both populations (Murray and Jackson, 1992). As such, the product of the concentrations of viruses and bacteria is directly related to the contact rate (Wilcox and Fuhrman, 1994). When this product is considered for the total cyanobacterial abundance multiplied by the sum abundance of the three measured cyanophage populations, an interesting pattern emerges; this product in GL 2013, RL 2012 and RL 2013 is statistically similar, around 10^6, while the product in GL 2012 is significantly lower with a mean around 10^3 (Pairwise comparisons with pooled SD, t-test $p<0.0005$). As the populations of cyanobacteria and cyanophages in were most correlated in GL 2012, it is possible that at low contact rates, covariance is more likely.

Beyond GL 2012, there are no strong relationships between cyanobacterial and cyanophage populations, suggesting that they are largely uncoupled. When CCA was employed to examine how the structure of one population influenced the structure of the other, no strong trends were observed, but results indicate that cyanobacterial populations had a greater influence on cyanophage populations than cyanophage had on cyanobacteria. The cyanobacteria explained 37% of variability within the cyanophage population while cyanophage explained only 8% of cyanobacterial variability (Table 2.4). Whereas no cyanophage significantly contributed to cyanobacterial variability, two out of the four cyanobacterial types significantly contributed to cyanophage variability. This suggests that cyanobacterial abundance is controlling the abundance

of cyanomyovirus populations examined in this study, and not the other way around. Previous studies examining cyanobacterial and cyanophage populations *in situ* observed variable patterns. One investigation found evidence that cyanophage were controlling cyanobacterial populations (Mühling et al., 2005), while another investigation observed the co-occurrence of cyanobacterial and cyanophage groups with no predator-prey dynamic (Sandaa and Larsen, 2006). The latter study observed seasonal trends similar to this investigation where cyanophage increased both before and after peaks in cyanobacterial abundance at different points during the study. Differences in the cyanophage populations examined between Mühling (2005), Sandaa and Larsen (2006) and this study may explain differing results and interpretations. Mühling (2005) examined cyanophage based on infection assays and thus could draw from a greater diversity of phage types with varying infection strategies, both this investigation and Sandaa and Larsen (2006) exclusively use molecular techniques to identify both cyanobacterial and cyanophage populations. Another possible difference is that both the current study and Sandaa and Larsen (2006) examine cyanomyoviruses belonging to the T4-like phage family. A recent study examining the dynamics of T4-like virus populations and bacteria over a three-year monthly time series at the San Pedro Ocean Time Series found that T4-like viruses often followed bacterial host abundance, suggesting that for this virus family, their abundance follows rather than controls their host *in situ* (Chow et al., 2013).

Shared and unique physical and chemical influences on cyanobacterial and cyanophage populations

Temperature and pH were factors that best explained the variability for both cyanobacterial and cyanophage populations. Population changes because of temperature suggest that cyanobacterial and cyanophage populations exhibit seasonal trends, which indeed appears to be the case based on observations of abundance over the two sampling periods (Figure 2.4). pH has been previously observed to contribute to the variability in bacterioplankton community structures

across freshwater lakes (Lindström et al., 2005). In GL and RL, pH varied from 7.6 to 8 over the time series with lower values appearing in the early spring.

Phage-exclusive factors are nitrate, bacterial abundance and total monthly precipitation. Although cyanophage dynamics are similar between the two lakes, cyanophage population dynamics appear variable between years. The cyanophage population exhibited a distinct shift in structure at one point during the 2012 sampling season, but the population exhibited a seasonal pattern in population structure in 2013 (Figure 2.4). The peak in cyanophage abundance in July of 2013 in both lakes appears to coincide with the regional precipitation pattern from that year (Figure 2.1A). A recent study examining the change in viral community structure after several heavy precipitation events found that viral abundance changed over the course of and after precipitation, and in one case, the precipitation led to a change that was reflected in both the viral and bacterial community structure (Williamson et al., 2014). The authors introduced three possible reasons for the increase in viral abundance and change in community structure in response to heavy precipitation: a) viruses are introduced into the lake via runoff, b) runoff introduces inducing agents into the lake that stimulate temperate phage induction and/or c) the perturbation caused by rainfall introduces sediment viruses back into the water column. In the case of the monitored GL and RL cyanophage populations, because these viruses are observed seasonally and are correlated with cyanobacterial assemblage structure, option "a" can be ruled out. The runoff may have introduced inducing agents into the lake, but even if this were the case, T4-like viruses are known to be lytic (Sullivan et al., 2005). Therefore, the change in community structure of this particular population of viruses is unlikely (but not impossible) because of increased induction of these phage populations. Finally, because GL and RL are meromictic, viruses were unlikely introduced from benthic sediments. However, heavy rainfall may have led to mixing within the mixolimnion that may introduce higher abundances of viruses from deeper populations of cyanophage into the surface waters. Another possibility is that viruses were resuspended from shallow sediments of the

lake's shelves that extend into the lake before the basin's steep decline. A previous study detected a cyanophage gene within shallow lake sediments of GL using qPCR (Hewson et al., 2012).

A relationship was noted between cyanophage populations between lakes (Mantel test, $r=0.7711$, $p=0.001$), however, no relationship was found when comparing cyanobacteria between lakes. The difference in between-lake patterns of cyanobacterial populations compared to cyanophage populations suggests that cyanobacterial population dynamics are driven by factors that differ between lakes that are not affecting the dynamics of the cyanophage population. SRP was the only variable that contributed to cyanobacterial variability that did not also contribute to cyanophage variability although it was the constraining variable of least significance for the cyanobacterial CCA (permutation test of significance, $p<1$, Table 2.4). Other ecological factors not measured in the current study may also be responsible for these differences. Previous studies have found differences in bacterioplankton community structures despite close proximity and similarity of aquatic habitats (Van Der Gucht et al., 2001; Eiler and Bertilsson, 2004; Murray et al., 1996). Biomanipulation of the fish population, changes in inflow (Van Der Gucht et al., 2001) and grazing (Tijdens et al., 2008) are all factors that could influence cyanobacterial community structure. Differences in the food-web dynamics of GL and RL are possible; the population of zooplankton within GL is hypothesized to feed on picocyanobacteria (Thompson and Ferris, 1990), although no previous study has measured the dynamics of higher trophic levels in RL.

Methodological Considerations

This study sampled a near-shore location in each of Green Lake and Round Lake periodically throughout two summer seasons. The use of one location, two meters off the shelf and 10cm below the surface is expected to be representative of the surface water picocyanobacterial population throughout the lake. Previous studies examining the dynamics of cyanobacteria at several different depths within GL found similar abundances at 0 and 4m depth (Thompson et al.,

1997), suggesting that the measurements of cyanobacteria and cyanophage in this study are representative of the upper mixolimnion.

In this study, the abundance of three lake-specific cyanophage was determined by the quantification of the T4-cyanophage portal protein using Taqman qPCR. Previous studies have considered only cyanophage populations capable of infecting a specific host (Mühling et al., 2005), or general populations of cyanophage using SYBR-based qPCR assays. Instead, this study uses a robust and reliable method to target very specific gene sequences in the environment (Short and Short, 2009). The g20 amplicons targeted in this study were between 83 and 141 bp in length. Between the primers and internal probe, 60 to 65 bp of the amplicon (42-72% of the amplicon sequences) were specifically targeted, and only those sequences with an exact match to the internal probe were quantified. Care was taken to insure specificity; this study used only primer sequences whose only hits via BLASTn against NCBI's nt database were to cyanophage sequences, or if the percent identity to the closest non-cyanophage match was <75%.

The presence and relative representation of different cyanobacterial types was monitored using cyanobacteria-specific ARISA (cARISA). Because of the large amplicon and standard sizes considered, improper identification of standard peaks was frequent using the automated analyses within the Peakscanner Software (ABI), as previously reported (Hewson and Fuhrman, 2006) . A number of additional smaller peaks were present within ARISA electropherograms. Cloning and sequencing of these peaks showed that they did not belong to picocyanobacteria, and thus they were omitted from the analysis. Relative representation of each picocyanobacterial phylotype was then converted into a semi-quantitative representation of the absolute abundance of each phylotype by multiplying the picocyanobacterial abundance at each time point by the relative representation of each peak at each time point. A previous study that investigated the contribution of peak height within ARISA profiles of *Prochlorococcus* in the marine environment found that relative peak area showed a consistent trend when compared to overall *Prochlorococcus* abundance

(Brown et al., 2005). In the current study, many steps were taken to insure the most consistent results between samples. Duplicate PCR reactions were pooled for each sample, duplicate ARISA runs were conducted for each sample, and the relative abundance of each phylotype was determined by dividing the sum peak area of each cyanobacterial phylotype for all electropherograms at each time point by the total peak area of only peaks determined to be cyanobacterial at that time point. Using this approach, this study considers the dynamics of the picocyanobacterial population, rather than considering only the presence or absence of different cyanobacterial phylotypes.

2.6 **Conclusions**

This study considers cyanobacterial and cyanophage population dynamics in two similar, connected lakes. Because of the similar characteristics of the lakes, the lakes were expected to experience similar phage-host dynamics. Surprisingly, this was not the case. Although cyanophage populations between the two lakes appeared highly correlated, the dynamics of four cyanobacterial phylotypes varied between lakes and years. Regardless, cyanobacterial and cyanophage population dynamics were only weakly correlated (Mantel test comparing cyanobacterial and cyanophage community distance matrices) and subsequent analyses showed that the populations are largely uncoupled. This study shows that even in highly similar freshwater aquatic systems, phage and host dynamics are variable annually and differ even between highly related systems. Future studies of phage-host dynamics in aquatic habitats must take a highly in-depth approach to understand factors contributing to this variable relationship to determine how this variability influences viral effects on lake biogeochemistry.

2.7 References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

Brown, M.V., Schwalbach, M.S., Hewson, I., and Fuhrman, J.A. (2005). Coupling 16S‐ITS rDNA clone libraries and automated ribosomal intergenic spacer analysis to show marine microbial diversity: development and application to a time series. Environ. Microbiol. *7*, 1466–1479.

Brunskill, G.J. (1968). Fayetteville Green Lake, New York: Physical & chemical limnology. Precipitation & sedimentation of calcite in a meromictic lake with laminated sediments. I.. II. (Cornell Univ.).

Brunskill, G., and Ludlam, S. (1969). Fayetteville Green Lake, New York. I. Physical and chemical limnology. Limnol Ocean. *14*, 817–829.

Callieri, C. (2008). Picophytoplankton in freshwater ecosystems: the importance of small-sized phototrophs. Freshw. Rev. *1*, 1–28.

Callieri, C., and Stockner, J.G. (2002). Freshwater autotrophic picoplankton: a review. J. Limnol. *61*, 1–14.

Cassman, N., Prieto‐Davó, A., Walsh, K., Silva, G.G., Angly, F., Akhter, S., Barott, K., Busch, J., McDole, T., and Haggerty, J.M. (2012). Oxygen minimum zones harbour novel viral communities with low diversity. Environ. Microbiol. *14*, 3043–3065.

Chamberlain, S.D., Kaplan, K.A., Modanu, M., Sirianni, K.M., Annandale, S., and Hewson, I. (2014). Biogeography of planktonic and benthic cyanobacteria in coastal waters of the Big Island, Hawai'i. FEMS Microbiol. Ecol.

Chow, C.-E.T., Kim, D.Y., Sachdeva, R., Caron, D.A., and Fuhrman, J.A. (2013). Top-down controls on bacterial community structure: microbial network analysis of bacteria, T4-like viruses and protists. ISME J. *8*, 816–829.

Clasen, J.L., Brigden, S.M., Payet, J.P., and Suttle, C.A. (2008). Evidence that viral abundance across oceans and lakes is driven by different biological factors. Freshw. Biol. *53*, 1090–1100.

Clokie, M.R., and Mann, N.H. (2006). Marine cyanophages and light. Environ. Microbiol. *8*, 2074–2082.

Crosbie, N.D., Pöckl, M., and Weisse, T. (2003). Dispersal and phylogenetic diversity of nonmarine picocyanobacteria, inferred from 16S rRNA gene and cpcBA-intergenic spacer sequence analyses. Appl. Environ. Microbiol. *69*, 5716–5721.

Danovaro, R., and Serresi, M. (2000). Viral density and virus-to-bacterium ratio in deep-sea sediments of the Eastern Mediterranean. Appl. Environ. Microbiol. *66*, 1857–1861.

Danovaro, R., Dell'Anno, A., Corinaldesi, C., Magagnini, M., Noble, R., Tamburini, C., and Weinbauer, M. (2008). Major viral impact on the functioning of benthic deep-sea ecosystems. Nature *454*, 1084–1087.

Van Der Gucht, K., Sabbe, K., De Meester, L., Vloemans, N., Zwart, G., Gillis, M., and Vyverman, W. (2001). Contrasting bacterioplankton community composition and seasonal dynamics in two neighbouring hypertrophic freshwater lakes. Environ. Microbiol. *3*, 680–690.

Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J.M., Furlan, M., Desnues, C., Haynes, M., Li, L., et al. (2008). Functional metagenomic profiling of nine biomes. Nature *452*, 629–632.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

Eiler, A., and Bertilsson, S. (2004). Composition of freshwater bacterial communities associated with cyanobacterial blooms in four Swedish lakes. Environ. Microbiol. *6*, 1228–1243.

Ernst, A., Becker, S., Wollenzien, U.I., and Postius, C. (2003). Ecosystem-dependent adaptive radiations of picocyanobacteria inferred from 16S rRNA and ITS-1 sequence analysis. Microbiology *149*, 217–228.

Fahnenstiel, G.L., and Carrick, H.J. (1992). Phototrophic picoplankton in Lakes Huron and Michigan: abundance, distribution, composition, and contribution to biomass and production. Can. J. Fish. Aquat. Sci. *49*, 379–388.

Fisher, M.M., and Triplett, E.W. (1999). Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. Appl. Environ. Microbiol. *65*, 4630–4636.

Hardbower, D.M., Dolman, J.L., Glasner, D.R., Kendra, J.A., and Williamson, K.E. (2012). Optimization of viral profiling approaches reveals strong links between viral and bacterial communities in a eutrophic freshwater lake. Aquat. Microb. Ecol. *67*, 59–76.

Haverkamp, T., Acinas, S.G., Doeleman, M., Stomp, M., Huisman, J., and Stal, L.J. (2008). Diversity and phylogeny of Baltic Sea picocyanobacteria inferred from their ITS and phycobiliprotein operons. Environ. Microbiol. *10*, 174–188.

Hewson, I., and Fuhrman, J.A. (2006). Improved strategy for comparing microbial assemblage fingerprints. Microb. Ecol. *51*, 147–153.

Hewson, I., Winget, D.M., Williamson, K.E., Fuhrman, J.A., and Wommack, K.E. (2006). Viral and bacterial assemblage covariance in oligotrophic waters of the West Florida Shelf (Gulf of Mexico). J. Mar. Biol. Assoc. U. K. *86*, 591–603.

Hewson, I., Barbosa, J.G., Brown, J.M., Donelan, R.P., Eaglesham, J.B., Eggleston, E.M., and LaBarre, B.A. (2012). Temporal dynamics and decay of putatively allochthonous and autochthonous viral genotypes in contrasting freshwater lakes. Appl. Environ. Microbiol. *78*, 6583–6591.

Hilfinger IV, M.F., Mullins, H.T., Burnett, A., and Kirby, M.E. (2001). A 2500 year sediment record from Fayetteville Green Lake, New York: evidence for anthropogenic impacts and historic isotope shift. J. Paleolimnol. *26*, 293–305.

Kumar, S., Nei, M., Dudley, J., and Tamura, K. (2008). MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief. Bioinform. *9*, 299–306.

Lindell, D., Jaffe, J.D., Coleman, M.L., Futschik, M.E., Axmann, I.M., Rector, T., Kettler, G., Sullivan, M.B., Steen, R., and Hess, W.R. (2007). Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. Nature *449*, 83–86.

Lindström, E.S., Kamst-Van Agterveld, M.P., and Zwart, G. (2005). Distribution of typical freshwater bacterial groups is associated with pH, temperature, and lake water retention time. Appl. Environ. Microbiol. *71*, 8201–8206.

Lu, J., Chen, F., and Hodson, R.E. (2001). Distribution, isolation, host specificity, and diversity of cyanophages infecting marine Synechococcus spp. in river estuaries. Appl. Environ. Microbiol. *67*, 3285–3290.

Lymer, D., and Lindström, E.S. (2010). Changing phosphorus concentration and subsequent prophage induction alter composition of a freshwater viral assemblage. Freshw. Biol. *55*, 1984–1996.

Lymer, D., Logue, J.B., Brussaard, C.P., BAUDOUX, A., Vrede, K., and LINDSTRÖM, E.S. (2008). Temporal variation in freshwater viral and bacterial community composition. Freshw. Biol. *53*, 1163–1175.

Madan, N.J., Marshall, W.A., and LAYBOURN‐PARRY, J. (2005). Virus and microbial loop dynamics over an annual cycle in three contrasting Antarctic lakes. Freshw. Biol. *50*, 1291–1300.

Mühling, M., Fuller, N.J., Millard, A., Somerfield, P.J., Marie, D., Wilson, W.H., Scanlan, D.J., Post, A.F., Joint, I., and Mann, N.H. (2005). Genetic diversity of marine Synechococcus and co‐occurring cyanophage communities: evidence for viral control of phytoplankton. Environ. Microbiol. *7*, 499–508.

Murray, A.G., and Jackson, G.A. (1992). Viral dynamics: a model of the effects of size, shape, motion and abundance of single-celled planktonic organisms and other particles. Mar. Ecol. Prog. Ser. Oldendorf *89*, 103–116.

Murray, A.E., Hollibaugh, J.T., and Orrego, C. (1996). Phylogenetic compositions of bacterioplankton from two California estuaries compared by denaturing gradient gel electrophoresis of 16S rDNA fragments. Appl. Environ. Microbiol. *62*, 2676–2680.

NOAA National Climatic Data Center (NCDC) | The world's largest active archive of weather and climate data producing and supplying data and publications for the world.

Noble, R.T. (2001). Enumeration of viruses. Methods Microbiol. *30*, 43–51.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M.H.H., Oksanen, M.J., and Suggests, M. (2007). The vegan package. Community Ecol. Package.

Olson, R.J., Chisholm, S.W., Zettler, E.R., and Armbrust, E.V. (1990). Pigments, size, and distribution of Synechococcus in the North Atlantic and Pacific Oceans. Limnol. Oceanogr. *35*, 45–58.

Parada, V., Sintes, E., van Aken, H.M., Weinbauer, M.G., and Herndl, G.J. (2007). Viral abundance, decay, and diversity in the meso-and bathypelagic waters of the North Atlantic. Appl. Environ. Microbiol. *73*, 4429–4438.

Patel, A., Noble, R.T., Steele, J.A., Schwalbach, M.S., Hewson, I., and Fuhrman, J.A. (2007). Virus and prokaryote enumeration from planktonic aquatic environments by epifluorescence microscopy with SYBR Green I. Nat. Protoc. *2*, 269–276.

Personnic, S., Domaizon, I., Dorigo, U., Berdjeb, L., and Jacquet, S. (2009). Seasonal and spatial variability of virio-, bacterio-, and picophytoplanktonic abundances in three peri-alpine lakes. Hydrobiologia *627*, 99–116.

Ranjard, L., Poly, F., Lata, J.-C., Mougel, C., Thioulouse, J., and Nazaret, S. (2001). Characterization of bacterial and fungal soil communities by automated ribosomal intergenic spacer analysis fingerprints: biological and methodological variability. Appl. Environ. Microbiol. *67*, 4479–4487.

Ripley, B. (2011). MASS: support functions and datasets for Venables and Ripley's MASS. R Package Version 7–3.

Rocap, G., Distel, D.L., Waterbury, J.B., and Chisholm, S.W. (2002). Resolution of Prochlorococcus and Synechococcus ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. Appl. Environ. Microbiol. *68*, 1180–1191.

Sandaa, R.-A., and Larsen, A. (2006). Seasonal variations in virus-host populations in Norwegian coastal waters: focusing on the cyanophage community infecting marine Synechococcus spp. Appl. Environ. Microbiol. *72*, 4610–4618.

Scanlan, D.J., Ostrowski, M., Mazard, S., Dufresne, A., Garczarek, L., Hess, W.R., Post, A.F., Hagemann, M., Paulsen, I., and Partensky, F. (2009). Ecological genomics of marine picocyanobacteria. Microbiol. Mol. Biol. Rev. *73*, 249–299.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., and Robinson, C.J. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. *75*, 7537–7541.

Schultze-Lam, S., Harauz, G., and Beveridge, T. (1992). Participation of a cyanobacterial S layer in fine-grain mineral formation. J. Bacteriol. *174*, 7971–7981.

Short, S.M., and Short, C.M. (2009). Quantitative PCR reveals transient and persistent algal viruses in Lake Ontario, Canada. Environ. Microbiol. *11*, 2639–2648.

Siering, P., Wolfe, G., Wilson, M., Yip, A., Carey, C., Wardman, C., Shapiro, R., Stedman, K., Kyle, J., and Yuan, T. (2013). Microbial biogeochemistry of Boiling Springs Lake: a physically dynamic, oligotrophic, low‐pH geothermal ecosystem. Geobiology *11*, 356–376.

Sullivan, M.B., Waterbury, J.B., and Chisholm, S.W. (2003). Cyanophages infecting the oceanic cyanobacterium Prochlorococcus. Nature *424*, 1047–1051.

Sullivan, M.B., Coleman, M.L., Weigele, P., Rohwer, F., and Chisholm, S.W. (2005). Three Prochlorococcus cyanophage genomes: signature features and ecological interpretations. PLoS Biol. *3*, e144.

Sullivan, M.B., Coleman, M.L., Quinlivan, V., Rosenkrantz, J.E., DeFrancesco, A.S., Tan, G., Fu, R., Lee, J.A., Waterbury, J.B., and Bielawski, J.P. (2008). Portal protein diversity and phage ecology. Environ. Microbiol. *10*, 2810–2823.

Suttle, C.A. (2007). Marine viruses—major players in the global ecosystem. Nat. Rev. Microbiol. *5*, 801–812.

Suttle, C.A., and Chan, A.M. (1994). Dynamics and distribution of cyanophages and their effect on marine Synechococcus spp. Appl. Environ. Microbiol. *60*, 3167–3174.

Takahashi, T., Broecker, W., Li, Y.H., and Thurber, D. (1968). Chemical and isotopic balances for a meromictic lake. Limnol. Oceanogr. *13*, 272–292.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. *30*, 2725–2729.

Team, R.C. (2012). R: A language and environment for statistical computing.

Thompson, J., and Ferris, F. (1990). Cyanobacterial precipitation of gypsum, calcite, and magnesite from natural alkaline lake water. Geology *18*, 995–998.

Thompson, J.B., Schultze-Lam, S., Beveridge, T.J., and Des Marais, D.J. (1997). Whiting events: biogenic origin due to the photosynthetic activity of cyanobacterial picoplankton. Limnol. Oceanogr. *42*, 133–141.

Thompson, L.R., Zeng, Q., Kelly, L., Huang, K.H., Singer, A.U., Stubbe, J., and Chisholm, S.W. (2011). Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. Proc. Natl. Acad. Sci. *108*, E757–E764.

Tijdens, M., Hoogveld, H.L., Kamst-van Agterveld, M.P., Simis, S.G., Baudoux, A.-C., Laanbroek, H.J., and Gons, H.J. (2008). Population dynamics and diversity of viruses, bacteria and phytoplankton in a shallow eutrophic lake. Microb. Ecol. *56*, 29–42.

Torgersen, T., Hammond, D., Clarke, W., and Peng, T. (1981). Fayetteville, Green Lake, New York: 3H-3He water mass ages and secondary chemical structure. Limnol Ocean. *26*, 110–122.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., and Leunissen, J.A. (2007). Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res. *35*, W71–W74.

U. S. Environmental Protection Agency (1983). Phosphorus, All Forms. Method 365.1 (Colorimetric, Automated, Ascorbic Acid). In In Methods for Chemical Analysis of Water and Wastes, (Cincinnati, OH, USA: U.S.E.P.A), pp. 365–1.1–365 – 1.7.

Waterbury, J.B., Watson, S.W., Guillard, R.R., and Brand, L.E. (1979). Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. Nature *277*, 293–294.

Weinbauer, M.G. (2004). Ecology of prokaryotic viruses. FEMS Microbiol. Rev. *28*, 127–181.

Wilcox, R.M., and Fuhrman, J.A. (1994). Bacterial viruses in coastal seawater: lytic rather than lysogenic production. Mar. Ecol.-Prog. Ser. *114*, 35–35.

Wilhelm, M.B., and Hewson, I. (2012). Characterization of Thrombolitic Bioherm Cyanobacterial Assemblages in a Meromictic Marl Lake (Fayetteville Green Lake, New York). Geomicrobiol. J. *29*, 727–732.

Wilhelm, S.W., Carberry, M.J., Eldridge, M.L., Poorvin, L., Saxton, M.A., and Doblin, M.A. (2006). Marine and freshwater cyanophages in a Laurentian Great Lake: evidence from infectivity assays and molecular analyses of g20 genes. Appl. Environ. Microbiol. *72*, 4957–4963.

Williamson, K.E., Harris, J.V., Green, J.C., Rahman, F., and Chambers, R.M. (2014). Stormwater runoff drives viral community composition changes in inland freshwaters. Front. Microbiol. *5*.

Wu, Q., Xing, P., and Liu, W.-T. (2010). East Tibetan Lakes Harbour Novel Clusters of Picocyanobacteria as Inferred from the 16S–23S rRNA Internal Transcribed Spacer Sequences. Microb. Ecol. *59*, 614–622.

Zhong, X., and Jacquet, S. (2013). Prevalence of viral photosynthetic and capsid protein genes from cyanophages in two large and deep perialpine lakes. Appl. Environ. Microbiol. *79*, 7169–7178.

Zhong, X., Berdjeb, L., and Jacquet, S. (2013). Temporal dynamics and structure of picocyanobacteria and cyanomyoviruses in two large and deep peri‐alpine lakes. FEMS Microbiol. Ecol. *86*, 312–326.

Chapter 3. Characterization of *Trichodesmium*- associated viral communities in the eastern Gulf of Mexico

Published in FEMS Microbiology Ecology, 6/1/2013, 84(3), p. 603-613

3.1 Abstract

Trichodesmium surface aggregations shape the co-occurring microbial community by providing organic carbon and nitrogen and surfaces on which microorganisms can aggregate. Rapid collapse of *Trichodesmium* aggregations leads to drastic changes in the chemical and physical properties of surrounding waters, eliciting a response from the microbial community and their associated viruses. Three viral metagenomes were constructed from experimentally lysed *Trichodesmium* collected from two locations in the eastern Gulf of Mexico. *Trichodesmium* were either treated with mitomycin C to induce potential lysogens, or incubated in the absence of mitomycin C. Comparative analyses of viral contiguous sequences indicated that viral composition was responsive to treatment type. Cyanophages were more represented within incubations treated with mitomycin C, while gammaproteobacterial phages were more represented within the untreated incubation. The detection of latent bacteriophage integrases in both the chemically treated and untreated incubations suggests that *Trichodesmium* death may lead to prophage induction within associated microorganisms. While no single cyanophage-like genotype associated with *Trichdesmium* lysis could be identified that might point to an infectious *Trichodesmium* phage, reads resembling *Trichodesmium* were recovered. These data reveal a diverse consortium of lytic and temperate phages associated with *Trichodesmium* whose patterns of representation within treated and untreated libraries offer insights into the activities of host and viral communities during *Trichodesmium* aggregation collapse.

3.2 Introduction

The marine diazotrophic cyanobacterium *Trichodesmium* spp. is an important component of tropical and sub-tropical oligotrophic marine ecosystems as a key source of nutrients and as a physical substrate for co-occurring microorganisms (Capone et al., 1997). *Trichodesmium* colonies support a wide diversity of associated microorganisms that likely consume fixed C and N that the cyanobacterium excretes as organic carbon and nitrogen. In addition, *Trichodesmium* provides a "pseudobenthic" habitat on which other organisms may colonize and grow (Sheridan et al., 2002; Hewson et al., 2009). In calm weather conditions, *Trichodesmium* colonies form dense surface aggregations that provide hot spots of biogeochemical cycling in oligotrophic surface waters. Both the cyanobacterium itself and bacteria that colonize the surface of *Trichodesmium* filaments are capable of enhanced phosphorus uptake relative to surrounding seawater bacterioplankton which is a key adaptation for growth in phosphorus limited cyanobacterial aggregations (Dyhrman et al., 2006; Van Mooy et al., 2012).

Trichodesmium aggregations undergo sudden population crashes causing them to disappear in short periods of time in both culture and field settings (Ohki, 1999; Berman-Frank et al., 2004; Hewson et al., 2004; Rodier and Le Borgne, 2010). Although some copepods are capable of consuming *Trichodesmium* (O'Neil and Roman, 1992), predation is not believed to be the primary cause of large mortality events as some species of *Trichodesmium* are known to excrete feeding deterrents (Hawser et al., 1992; Capone et al., 1997). One hypothesis for *Trichodesmium* mortality is autocatalyzed programmed cell death. This has been observed in the marine phytoplankton *Emiliana huxleyi,* and was hypothesized to occur in *Trichodesmium* based on gene expression and translation of a catalase-like gene, associated with eukaryotic programmed cell death, during *Trichodesmium* demise (Berman-Frank et al., 2004, 2007). Simultaneous research has shown, however, that viral lysis is the cause of bloom collapse in many bloom forming marine algae (Nagasaki, 2008), and further research of *Emiliana huxleyi* bloom dynamics has revealed that a

virus contributes to coccolithophore demise, with host lysis due to viral infection occurring through a pathway similar to a programmed cell death pathway that occurs in multicellular eukaryotes (Vardi et al., 2009). Viral infection of *Trichodesmium* is another hypothesized cause of rapid collapse of *Trichodesmium* aggregations. Previous studies have observed virus-like particle production from *Trichodesmium* exposed to the mutagen mitomycin C (Ohki, 1999; Hewson et al., 2004), commonly used in studies of lysogeny in seawater (Weinbauer and Suttle, 1999; McDaniel et al., 2001), in both lab and field incubations of *Trichodesmium.* Thus, the emergence of virus-like particles upon mitomycin C treatment suggests that *Trichodesmium* harbor prophage that may enter the lytic cycle under as-yet undefined conditions (Ohki, 1999; Hewson et al., 2004).

Trichodesmium aggregation collapse inevitably leads to changes in the physical and chemical conditions of seawater that likely impact the associated microbial communities. Previous studies have found a wide diversity of organisms associated with *Trichodesmium* colonies and *Trichodesmium* aggregations. Bacterial abundance on *Trichodesmium* filaments is three-fold higher per unit volume than abundance in the surrounding water (O'Neil and Roman, 1992; Sheridan et al., 2002), and the taxonomic structure of communities associated with *Trichodesmium* is different than typical open ocean bacterioplankton (Hewson et al., 2009). Like algal blooms, collapse of surface aggregations of *Trichodesmium* causes large pulses of dissolved and particulate organic matter from dying *Trichodesmium* cells, leading to a shift in the sources of C and N available for co-occurring microorganisms (Rodier and Le Borgne, 2010). In addition, collapse and disappearance of *Trichodesmium* colonies reduces the availability of physical surfaces onto which associated microbes may aggregate. The large chemical and physical perturbations caused by *Trichodesmium* death may lead to a shift in the activity of associated microorganisms and their viruses.

The aims of the current study were twofold. Firstly, this study used high throughput sequencing of the virus fraction of *Trichodesmium* aggregations upon collapse to examine the taxonomic structure of the associated viral consortia. This study compared viruses present after

Trichodesmium aggregations were chemically treated (mitomycin C) to induce lysogenic viruses to the viral consortia present after containment-induced *Trichodesmium* lysis. The second goal was to investigate potential viruses of *Trichodesmium* within these incubations. Results of sequence analyses suggest that viruses capable of infecting many different phylogenetic groups were present in the treated and untreated incubations. Phages of bacterial groups previously found in association with *Trichodesmium* were detected and differences in relative representation of viral consortia between mitomycin C treated and untreated incubations suggested that viruses that arose in response to *Trichodesmium* lysis without chemical treatment were different than the viruses induced due to mitomycin C addition. The overall similarity between mitomycin C treated and untreated incubations, however, suggests that *Trichodesmium* demise may cause lysogenic induction even when no lysogen is added. No single viral genotype resembling known cyanophage made up a significant proportion of these libraries, and thus no definitive virus of *Trichodesmium* could be identified, leaving room for further inquiry into the cause of rapid *Trichodesmium* collapse.

3.3 Methods

Sample Collection and Experimental Setup

Experiments were conducted in the eastern Gulf of Mexico onboard the R/V Pelican cruise EK0001, 6 – 10 October 2009. *Trichodesmium* colonies were collected at 2 stations: The mouth of Tampa Bay (27°32' 50"N, 82° 46' 55"W); and station 9B (26°25'44" N, 82°30'58"W), both on the West Florida Shelf. These stations were chosen because dense *Trichodesmium* aggregations were visible at the water's surface. *Trichodesmium* were obtained using a 1 m diameter, 125 μm mesh plankton net which was towed behind the vessel for 5 mins to collect colonies. *Trichodesmium* collected by net tow were further concentrated over a 10 μm Isopore (PC, Millipore) filter and resuspended in 250 ml of 10μm - filtered seawater in acid-washed and seawater rinsed polycarbonate bottles. At the Tampa Bay station, collected water was split into two incubations and mitomycin C was added (20 ng ml^{-1}, final concentration) to one of the incubations, while the other incubation was left untreated. A third incubation with mitomycin C added was conducted at station 9B.

Trichodesmium were then incubated in flow-through outdoor incubators to maintain ambient surface water temperature, and subject to 25% attenuated irradiance under shade cloth for 12 hours. After 12 hours, *Trichodesmium* colonies in all incubations were green and the water pink, indicating the presence of cyanobacterial phycoerythrin in the water and thus *Trichodesmium* death. The entire incubation was filtered through 0.2 μm Durapore (PVDF; Millipore) filters and the filtrate was transferred to sterile 50 ml tubes, which were frozen in liquid nitrogen until further analysis. Filtration through a 0.2 μm filter is necessary to remove bacteria and other large debris, but may have biased the composition of our metaviromic datasets because some viruses are around this size threshold, such as some phycodnaviridae (Simpson et al., 2003) as well as long-tailed siphoviridae (Sullivan et al., 2009).

Viral Nucleic Acid Extraction and Preparation

Purification of viral particles followed a modified filter-chloroform-nuclease treatment protocol (Ng et al., 2011). The 0.2 µm-filtered *Trichodesmium* lysates were precipitated in 10% Polyethyleneglycol (PEG 8000, weight/volume) at 4°C for 24 hour. The PEG treated lysates were then centrifuged at 30,000 x*g* for 30 minutes. The supernatant was decanted and pellets were resuspended in 1 ml of 0.02 µm–filtered phosphate buffered saline (PBS). A subsample (100 µl) of each sample was filtered again through 0.2 µm Acrodisc (0.2 um; Pall Gellman, New York) filters to insure that samples were free of potential contaminating particles (like bacteria). Chloroform (0.2 volumes) was added to each sample, which were gently agitated and settled at room temperature for 10 min. Samples were briefly centrifuged and the aqueous layer (containing purified viruses) was transferred to a new tube. The aqueous fraction was incubated with DNase (2.5 U each) for 3 h at 37 degrees C, after which EDTA was added to a final concentration of 2 mM. Viral DNA in the purified aqueous layer was extracted using the ZR viral DNA extraction kit (Zymo Research, Irvine, CA) according to manufacturer's protocols.

Nucleic Acid Amplification and Sequencing

Extracted viral DNA was amplified via multiple displacement (Φ29) amplification in triplicate reactions (Genomiphi©, GE Biosciences, Piscataway, NJ). Triplicate reactions were combined and amplified DNA was purified using a modified DNeasy extraction protocol (Thurber et al., 2009). The purified and amplified DNA was subjected to high throughput sequencing using 454 Titanium chemistry at EnGenCore (University of South Carolina, Columbia, SC), where each sample was sequenced on 1/8th of a picotiter plate. Sequence data is available at the Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis (CAMERA) website under project accession number CAM_P_0000951.

Viral Metagenome Phylogenetic Annotation and Comparative Analyses

Annotation of viral metagenome reads and contiguous sequences followed an approach similar to VIROME (http://virome.diagcomputing.org/#view=Methods; Bhasvar et al., 2009). Viral metagenomic libraries were cleaned up using the QC filter available on the Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis (CAMERA) website (http://camera.calit2.net; Sun et al., 2011), and then de-replicated in order to account for 454 sequencing amplification bias (Gomez-Alvarez et al., 2009) using the 454 de-replication workflow also available at CAMERA. Sequence reads within each cleaned library were subjected to BLASTx (Altschul et al., 1990) against the viral proteins, prokaryotic proteins, and fungal proteins databases at CAMERA with an e-value cut-off of $e < 10^{-3}$. The top hit from all three databases based on lowest e-value (and highest bit score amongst annotations with tied e-values) were compiled into Microsoft Excel. The phylogenetic affiliation of viral hits was annotated based on the NCBI Taxonomy browser (http://www.ncbi.nlm.nih.gov/Taxonomy/). The host genus of viral hits was determined based upon the name of the virus (e.g. *Prochlorococcus* phage P-SSM4) or by a literature search to determine the organism of isolation, if the host organism is not described within the name of the phage (e.g. phage λ).

Sequence reads were assembled into contiguous sequences (contigs) in CLC Genomics Workbench 4.0 using stringent assembly parameters (20% minimum overlap, 95% identity). Open reading frames (ORFs) were identified on contigs using fraggene_scan (Rho et al., 2010). ORFs were compared using BLASTp against the Reference Proteins database (CAMERA) using an e-value cut-off of $e<10^{-3}$. Contigs were annotated based on their best ORF match to the CAMERA Reference Proteins database. Representation of individual viral genotypes was determined by taking the total contig base pairs of all contigs with best ORF matches to specific sequenced phages and dividing that number by the total number of basepairs in the assembled contig library for each dataset.

To compare assembled contiguous sequence annotations between libraries, the relative representation of viral groups was calculated based upon the sum of the lengths of contigs assigned to phylogenetic bins divided by the total length of all contigs within that particular analysis (e.g. viral contigs; Equation 1).

Equation 1: Difference between mitC-treated libraries and untreated libraries=

$$\frac{\sum bp\_mitC\_contigs\_within\_category}{\sum bp\_all\_mitC\_contigs\_within\_analysis} - \frac{\sum bp\_untreated\_contigs\_within\_category}{\sum bp\_all\_untreated\_contigs\_within\_analysis}$$

Read libraries were compared to each other using reciprocal BLASTn comparisons. In addition to the three libraries generated for this study the libraries were compared to the Induced Tampa Bay Phage library constructed using similar techniques (CAM_PROJ_TampaBayPhage; McDaniel et al., 2008) and viral reads from the Marine Viromes dataset available at CAMERA (CAM_PROJ_MarineVirome), divided by location (Sargasso Sea (SAR), Arctic (Arctic), British Columbia coastal waters (BBC) and the Gulf of Mexico (GOM); Angly et al., 2006). 10,000 reads were randomly sampled from each library and subjected to BLASTn against all other read libraries ($e<10^{-10}$), using each library as both a database and a query library for each reciprocal comparison. The number of reads in each query hitting at least one read in the database library was recorded, and the average number of reads in the query with hits to the database was calculated for each pair wise comparison. A similarity matrix was then constructed based upon the averaged results of three separate comparisons and subjected to multidimensional scaling (MDS) in XL Stat (Addinsoft SARL, New York, NY) to visualize the relatedness between libraries.

3.4 Results and Discussion

Library Statistics

Read libraries had between 30,648 and 69,208 reads, after de-replication and clean-up, representing 62 Mbp of sequence information. G+C content in all three libraries was below 50%. After cleanup, between 20 and 30% of reads in each library matched known viral, prokaryotic and eukaryotic fungal proteins based on BLASTx comparison to protein databases (e-value < 10^{-3}). Of annotated reads, one third to nearly one half of the reads most closely matched viral proteins, with a larger proportion matching bacterial proteins. Contiguous sequence (contig) assembly resulted in 2,263 to 2,687 contigs per library (Table 3.1). Average contig length was 755 - 875 bp and annotation revealed a similar proportion of viral, bacterial, archaeal and unknown contigs (Figure 3.1A).

Comparison Between Libraries and to Other Viral Datasets

The mitomycin C treated and untreated *Trichodesmium* viral metagenomes from Tampa Bay were most similar to each other, and the mitomycin C treated library from station 9B was more similar to the two Tampa Bay viral metagenomes than to other sequenced viral metagenomes included in our analysis (Figure 3.2). This suggests that the viral community associated with *Trichodesmium* lysis is distinct from virioplankton communities in other environments. In addition, the mitomycin C treated and untreated libraries from the mouth of Tampa Bay were most similar of any of the libraries in this analysis, with an average of 74% read similarity based on BLASTn with an e-value cut-off score of 10^{-10}.

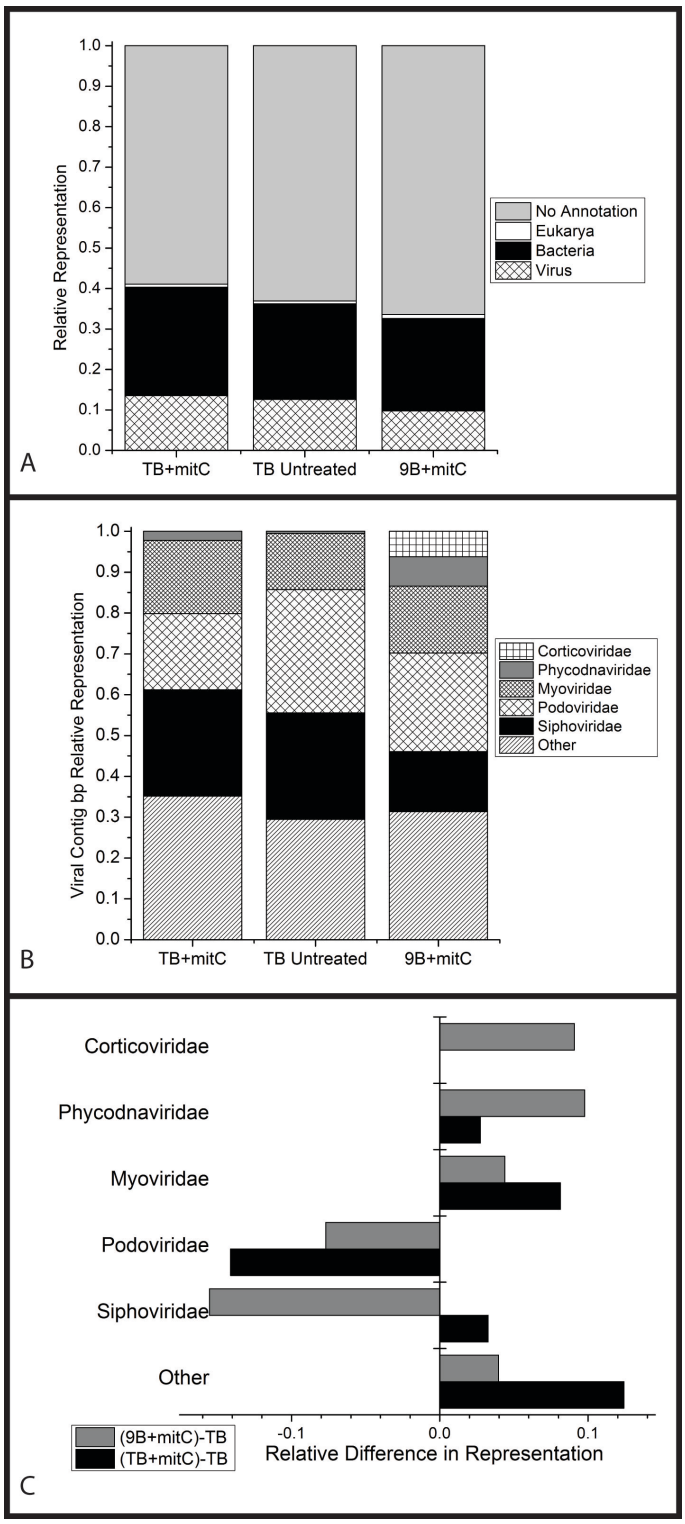


Figure 3.1. (A) Contiguous sequence annotations of kingdom based on best ORF match per contig to CAMERA Reference Proteins, $e<10^{-3}$, (B) relative representation of viral families amongst contigs with best ORF match to viral proteins normalized by contig base-pairs falling into each family divided by the total number of viral contig basepairs; (C) relative difference between mitomycin C treated and untreated Tampa Bay viral contigs (see equation 1). Values that fall below zero indicate decreased representation in a mitomycin C treated library compared to the untreated Tampa Bay library.

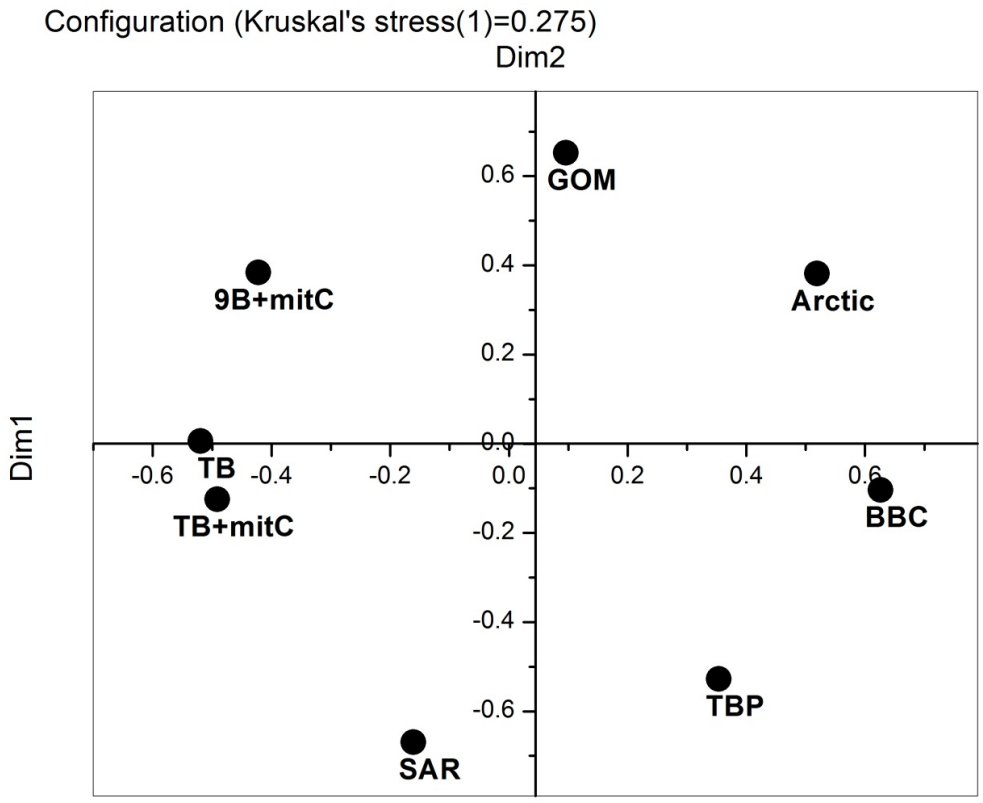


Figure 3.2. MDS (Multi-dimensional scaling) plot of read libraries based on a similarity matrix of averaged triplicate reciprocal BLASTn analyses of 10,000 random reads from each library. TB+mitC=Tampa Bay mitomycin c treated, TB=Tampa Bay untreated, 9B+mitC=9B mitomycin C treated, SAR=Marine Viromes Sargasso Sea subset, Arctic= Marine Viromes Arctic subset, BBC= Marine Viromes British Columbia coastal waters subset, GOM= Marine Viromes Gulf of Mexico Subset, TBP=Mitomycin C-treated Tampa Bay metavirome from McDaniel et al. 2008.

Assessing φ29 Biases

Analysis of sequence reads revealed a variable distribution of read hits, with a number of similar reads resembling specific genes present in very high levels. Because libraries were prepared using Φ29 amplification, there are expected biases in the representation of genes and organisms within individual libraries (Shoaib et al., 2008). Subsequent contig based analyses ruled out some potential biases associated with over-representation of genes relative to template quantities. For example, read-base representation of genes within the *Pseudalteromonas phage PM2* genome in the 9B mitomycin C treated read library was uneven, as some genes had greater than 100x coverage but adjacent genes had only 1x coverage. Stringent assembly of sequence reads into contigs

removed some over-amplification biases by combining many highly amplified reads onto single contigs. The length of contigs is used in this study as a proxy for representation of genotypes within libraries. Annotated contigs then represent the well represented viruses within each of the samples, and as these libraries were all constructed in the exact same fashion, we will use differences in the contig lengths assigned to different viral genotypes and virus categories between libraries to assess differences in viral community structure. The following comparisons are relative comparisons of the same viral genotypes and categories between libraries, and are not quantitative. In addition, read based annotations were used for characterization of individual targeted genes, and assembled contiguous sequence ORF based annotations were used for comparative analyses between libraries.

Table 3.1. Library statistics for reads and assembled contigs in the Tampa Bay mitomycin C-treated (TB+mitC) and untreated Tampa Bay (TB) and the 9B mitomycin C treated (9B+mitC) libraries. Contigs considered viral had best ORF match (based on BLASTp, lowest e-value and highest bit score) to the Reference Proteins database at CAMERA.

	TB+mitC	TB	9B+mitC
# Raw Reads	65,832	79,407	34,358
Average Raw Read Length	326 ± 0.80	355 ± 0.45	374 ± 0.84
# Reads	57,493	69,208	30,648
Total base pairs (bp)	24,375,865	25,581,060	12,478,835
Average % GC	45 ± 0.02	45 ± 0.02	43 ± 0.04
Average Read Length	424 ± 0.51	370 ± 0.44	407 ± 0.70
# Contigs	2,516	2,687	2,263
Average Contig Length	856 ± 18	826 ± 22	755 ± 13
Total Contig bp	2,152,671	2,220,060	1,709,318
# Viral Contigs	341	339	221
Total Viral Contig bp	316,481	350,470	184,625
Avg Viral Contig Length	942 ± 81	1110 ± 108	875 ± 56
Viral Contig Range	239 - 16870	191 - 15790	249 - 10988

Annotation of Contiguous Sequence ORFs

Approximately 10% of contigs from each library had best ORF matches to previously described viral proteins (Figure 3.1A), which represented a total of 215 viral genotypes based on the strongest ORF match per contig. Different phages exhibited different patterns of representation amongst the three libraries (Figure 3.3). Contigs with best ORF matches to a subset of viruses were more highly represented within the untreated Tampa Bay viral metagenome compared to the mitomycin C treated viral metagenomes. These included contigs annotated as *Pseudoalteromonas phage pYD6-A*, *Enterobacter phage EcP1*, *Pseudomonas phage LIT1*, *Vibrio phage VBP47* and *Erwinia phage vB_EamP-S6*. These phages all infect gammaproteobacteria, and have either unknown or lytic infection strategies. Another group of viruses exhibited increased representation in both the mitomycin C treated libraries compared to the untreated library. Some of the highest represented members of this group include *Methylophilales phage HIM624-A*, *Roseobacter phage SIO1*, *Cyanophage KBS-S-2A*, *Synechococcus phage S-SM2* and *Synechococcus phage S-CAM8*. The only member of this described group whose infection strategy is known is *Roseobacter phage SIO1*, which was originally characterized a lytic phage of *Roseobacter*. *Synechococcus* phages *S-SM2* and *S-CAM8* are both members of the myoviridae viral family. To date, there is little evidence that cyanobacterial myoviruses are capable of lysogeny (Clokie et al., 2010), thus their higher representation in the chemically induced libraries is interesting. A third group of viruses were most highly represented in one of the mitomycin C treated viral metagenomes but had lower representation within the other two libraries. Highly represented viral genotypes within the Tampa Bay mitomycin C treated dataset included *Salicola phage CGphi29*, *Vibrio phage pYD21-A*, *Pseudoalteromonas phage H105/1* and *Vibrio phage jenny 12G5*. Several of these phages are known to be temperate, suggesting that these groups were induced upon treatment with mitomycin C. *Pseudoalteromonas phage PM2* was most highly represented in the 9B mitomycin C treated library, and was absent from both libraries from Tampa Bay (Figure 3.3, Table 3.2).

Contigs with best ORF matches to members of the *Caudovirales*, or tailed phage, were the most highly represented viral order with the families *Podoviridae, Siphoviridae* and *Myoviridae* together comprising the majority of identifiable viral sequences in each library. Sequences matching the *Phycodnaviridae* had a consistent presence within all three contig libraries, with the largest representation within the mitomycin C treated 9B viral metagenome. Phycodnaviruses are large viruses whose sizes range from 0.15 to greater than 0.2 μm (Simpson et al., 2003) which is at or above the pore size of the 0.2 um pre-filter used. Thus the phycodnaviruses present within these datasets may not represent all phycodnaviruses present at the time of sampling. An 11,000 bp contig with several ORFs best resembling proteins of *Pseudoalteromonas phage* PM2, was only found in the mitomycin C-treated library collected at station 9B. There were differences between the relative representation of viral families within mitomycin C-treated and untreated Tampa Bay viral metagenomes (Figure 3.1B and 3.1C). Myoviruses were slightly overrepresented in the untreated library compared to both mitomycin C-treated libraries. There was little difference in representation of contigs matching siphoviruses between the Tampa Bay mitomycin C treated and untreated libraries, but there was a decrease in their representation in the mitomycin C-treated 9B viral metagenome. Finally, there was a greater representation of contigs matching Phycodnaviruses (most closely resembling viruses of *Chlorella*, *Ostreococcus* and *Micromonas*) within the mitomycin C-treated viral metagenomes compared to the untreated viral metagenome. The cause of enhanced representation of phycodnaviruses is unclear, since mitomycin C is not expected to induce viruses within eukaryotic organisms.

Table 3.2. 20 most highly represented viral genotypes within the contig libraries.

Virus	Taxon ID	Virus Family	Infection Strategy	Most Represented
Pseudoalteromonas phage pYD6-A	754052	Unclassified dsDNA virus	unknown	TB untreated
Salicola phage CGphi29	754067	Unclassified dsDNA virus	unknown	TB+mitC
Methylophilales phage HIM624-A	889949	Unclassified dsDNA virus	unknown	TB+mitC
Enterobacter phage EcP1	942016	Podovirus	unknown	TB untreated
Pseudomonas phage LIT1	655098	Podovirus	lytic	TB untreated
Roseobacter phage SIO1	136084	Podovirus	lytic	TB+mitC
Vibrio phage pYD21-A	754049	Unclassified dsDNA virus	unknown	TB+mitC
Vibrio phage douglas 12A4	573171	Unclassified	unknown	TB untreated
Vibrio phage VBP47	754073	Unclassified dsDNA virus	unknown	TB untreated
Erwinia phage vB_EamP-S6	1051675	Podovirus	lytic	TB untreated
Synechococcus phage S-SSM7	445686	Myovirus	lytic	TB+mitC
Pseudoalteromonas phage H105/1	877240	Siphovirus	temperate	TB+mitC
Pseudomonas phage YuA	462590	Siphovirus	temperate	TB+mitC
Cyanophage KBS-S-2A	889953	Unclassified dsDNA virus	unknown	9B+mitC
Pseudoalteromonas phage PM2	10661	Corticovirus	lytic	9B+mitC
Pseudomonas phage LUZ7	655097	Podovirus	lytic	TB untreated
Vibrio phage jenny 12G5	573176	Unclassified	unknown	TB+mitC
Synechococcus phage S-SM2	444860	Myovirus	unknown	9B+mitC
Synechococcus phage S-CAM8	754038	Myovirus	unknown	9B+mitC
Phage phiJL001	279383	Siphovirus	temperate	TB untreated

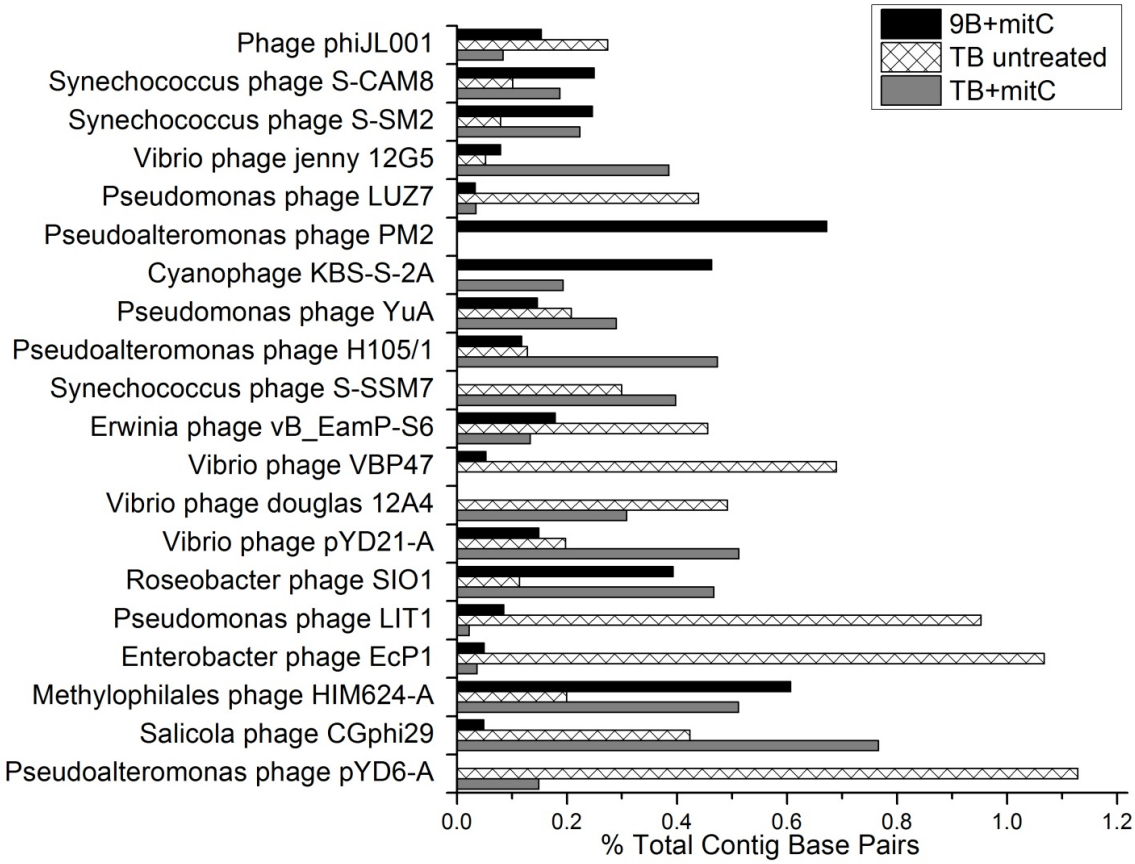


Figure 3.3. Relative representation of 20 most abundant viral contig annotations, annotated based on best ORF match per contig. Relative representation determined by total contig base pairs falling into each annotation divided by the total contig bp in each library.

Viruses of alphaproteobacteria, bacteroidetes, betaproteobacteria, cyanobacteria, firmicutes, and gammaproteobacteria were the six most dominant viral groups based on bacterial host. Previously described viruses of gammaproteobacteria were more represented within the untreated Tampa Bay viral metagenome compared to the mitomycin C-treated viral metagenomes. Conversely, cyanophage were better represented in the mitomycin C-treated viral metagenomes compared to the untreated viral metagenome (Figure 3.4). Together, these results indicate that lysogenic cyanobacteria were likely sensitive to mitomycin C addition. The large representation of gammaproteobacterial viruses in these libraries suggests that gammaproteobacteria were the most active and potentially the most abundant hosts. These observations are congruent with a previously reported *Trichodesmium* community metatranscriptomes which found abundant

gammaproteobacterial transcripts (Hewson et al., 2009). Phages of hosts that comprised a large proportion of metatranscriptomic reads in that study, particularly *Pseudoalteromonas* and *Alteromonas* were well represented within all three metaviromic libraries in this study. These organisms, some of whom are often referred to as "copiotrophs" (Ivars-Martinez et al., 2008), may have responded to the increased DOM availability after *Trichodesmium* lysis, which subsequently led to a response from the related phage population either through induction of prophage, or through enhanced lytic infection due to increases in host density. The lower representation of gammaproteobacterial phage genotypes within the two mitomycin C-treated contig libraries compared to the untreated library may indicate that this group waas not the dominant host at the onset of the incubation experiments, or that these hosts could not respond to *Trichodesmium* lysis in the presence of mitomycin C. Interestingly, there is almost no difference in the relative representation of alphaproteobacterial, betaproteobacterial and firmicute phage genotypes between the mitomycin C-treated and untreated libraries, suggesting that the representation of these host groups remains similar in both incubations, and that the viral population responds similarly to mitomycin C addition and *Trichodesmium* lysis (Figure 3.4).

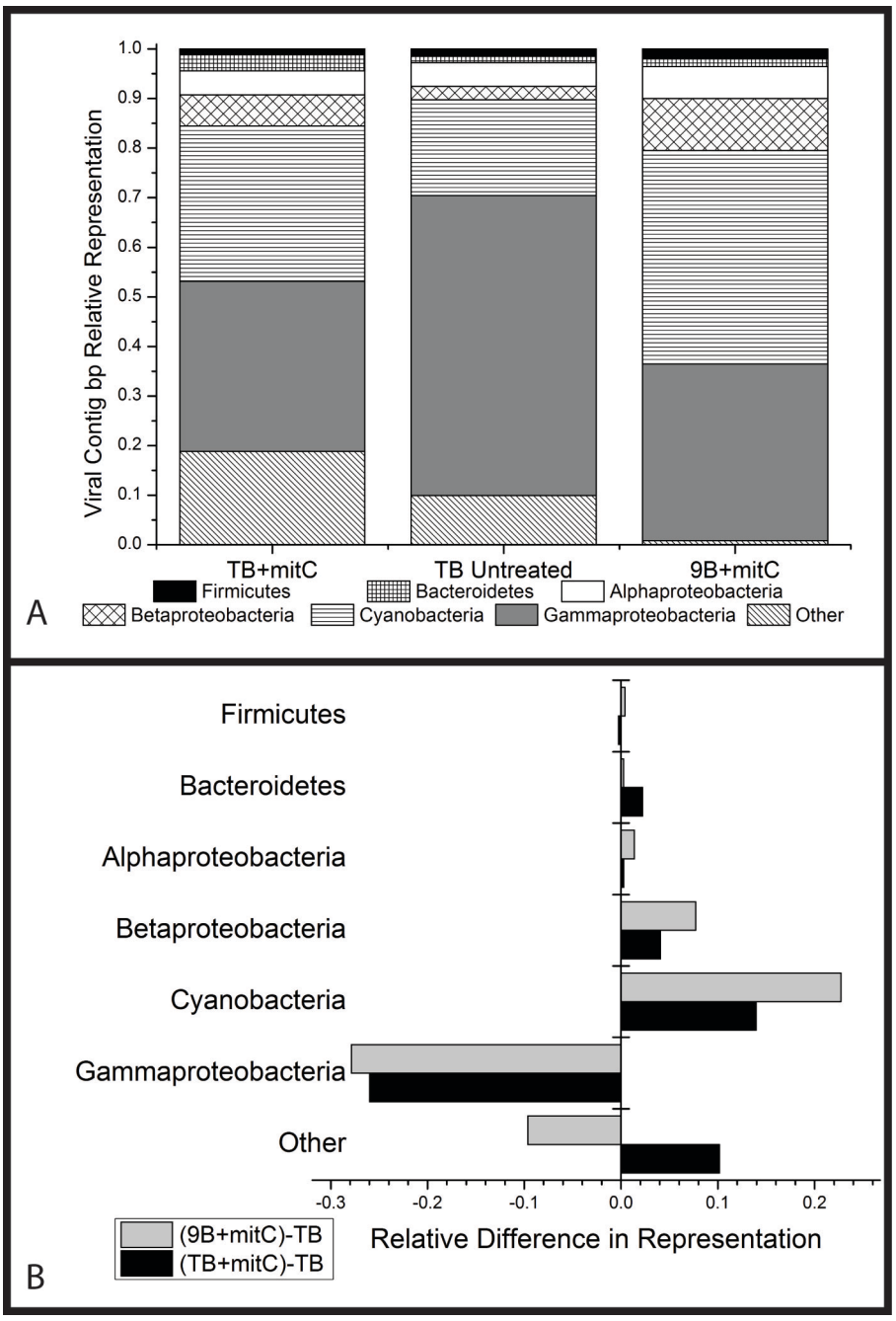


Figure 3.4. Relative representation of viral contigs based on host organism (A) and relative difference between mitomycin C treated and untreated Tampa Bay viral contigs (B).

Mitomycin C treatment and location are the two factors that are most likely responsible for the differences in viral composition in this study. All three viral metagenomes were generated from incubations in which *Trichodesmium* lysed, leading to enhanced input of dissolved organic matter

into the incubations, but in the two incubations in which mitomycin C was added, the associated bacterial community would not be able to respond to these physical changes before mitomycin C caused prophage induction. The phage genotypes that were better represented in the untreated viral metagenome compared to the mitomycin C-treated libraries represent viruses that may have entered a lytic infection cycle in response to changes in host activity, or lytic viruses that increased in abundance due to increases in host abundance. Lytic viral production typically occurs when host and free virus abundance is high (Wilcox and Fuhrman, 1994). Lytic viruses have disproportionate effects on dominant bacterial taxa in plankton due to the density-dependent and relatively host-specific dynamic between viruses and their hosts. Thus, the most metabolically dominant or fastest growing member of the microbial community is controlled most heavily by lytic viruses, allowing less competitive community members to co-exist, an interaction termed "kill the winner" (Thingstad, 2000). The viral genotypes more highly represented within the untreated incubation may have been interacting with hosts in a "kill-the-winner" scenario, whereby hosts of these viruses were the fastest growing portion of the bacterial community. Conversely, the viruses that were better represented in the two mitomycin C treated viral metagenomes likely represent temperate phages arising from prophage within hosts sensitive to mitomycin C that were in high abundance at the beginning of the incubation and viruses already present within the community at the time of mitomycin C addition. The latter description may explain why several cyanobacterial myoviruses, not expected to be lysogenic, were more represented within the mitomycin C treated incubation; associated picocyanobacteria and their phages were abundant members of the community before *Trichodesmium* lysis, and thus were represented within the mitomycin C treated incubations that contained a viral community that was not able to respond to *Trichodesmium* lysis. Finally, viral genotypes that exhibited higher representation in only one of the mitomycin C treated incubations likely represent variation in host and/or viral community composition between sampling locations. This may be due to differences in the physical and nutrient characteristics at the

two locations, or differences in *Trichodesmium* colony morphology between locations, which has been previously observed to lead to differences in the structure of associated microbial assemblages (O'Neil and Roman, 1992; Sheridan et al., 2002; Hmelo, 2010).

Integrase Genes Present within Read Libraries

Bacteriophage integrase genes have previously been used to characterize lysogeny within metaviromic samples (McDaniel et al., 2008). In this study, integrases made up 0.28-0.38% of sequence reads in each viral metagenome. Within the 9B mitomycin C treated library, integrases most closely matching *Vibrio* phage were most highly represented followed by those of *Pseudoalteromonas* phage. Phage integrases of two other marine gammaproteobacterial hosts (*Idiomarina* and *Congregibacter)* were detected, with equal representation of the two in the untreated library and a higher representation of the *Congregibacter* phage integrase in the mitomycin C treated viral metagenome. The representation of integrases in the Tampa Bay mitomycin C treated library was 0.1% higher than the representation of integrases within the untreated Tampa Bay library (Table 3.3). The large proportion of integrase-like reads resembling phage of gammaproteobacterial hosts (*Vibrio*-like, *Congregibacter*-like), suggests that prophage induction of these or similar gammaproteobacterial hosts was stimulated. Similarity of unique integrase-like sequences between untreated and mitomycin C treated Tampa Bay viral metagenomes suggests that temperate phages were induced after mitomycin C addition, but also by containment or *Trichodesmium* collapse.

Table 3.3. Summary of integrase reads found within the three libraries.

Integrase Summary

	TB+mitC	TB Untreated	9B+mitC
# Integrase	220	197	104
% Integrase Reads	0.38	0.28	0.34
Unique hits	43	45	55

Analysis of Reads Resembling Trichodesmium

Sequence reads with varying degrees of similarity to 19 different proteins of *Trichodesmium erythraeum IMS101* were found in the metaviromic libraries. *Trichodesmium* proteins detected included a phage integrase (in the untreated Tampa Bay viral metagenome only), 2 different Rho-termination factor-like proteins (in both mitomycin C treated and untreated Tampa Bay viral metagenomes), a RNA polymerase sigma factor (in 9B mitomycin C treated) and several hypothetical proteins. Interestingly, most reads matching *Trichodesmium* proteins were found in the untreated *Trichodesmium* viral metagenome, and the least in the 9B mitomycin C treated viral metagenome. Sequence reads matching 3 proteins were found in both mitomycin C treated and untreated Tampa Bay viral metagenomes: a DUF900 hydrolase-like protein (RefSeq ID YP_721790), a hypothetical protein (YP_723033), and a Rho-termination factor-like protein (YP_720117). Despite the large biomass of *Trichodesmium* in incubations, and apparent large viral burst size (Hewson et al., 2004), no single viral genotype that dominated libraries which would indicate a virus involved in *Trichodesmium* demise could be identified. It is possible that a phage of *Trichodesmium* was present at relatively low abundances within the phage community since other

phage, particularly those infecting taxa that were likely consuming organic matter of the dying cyanobacteria, comprised a large proportion of total viral abundances in these incubations. Another possibility is that there are multiple, diverse phage that infect *Trichodesmium*, which would also lead to inconclusive results based on the size of the libraries sequenced. The *Trichodesmium* integrase gene as well as the greatest number of other *Trichodesmium*-like sequence reads were present in the untreated library. Sequences of a potential *Trichodesmium* phage may have been pushed below the sequence coverage of our viral metagenomes. The consistent detection of the same *Trichodesmium* genes in the two Tampa Bay viral metagenomes suggests this approach was not biased towards sampling the genome of *Trichodesmium,* but rather repeated detection of phage particles bearing *Trichodesmium*-like genes. In addition to *Trichodesmium*-like genes, there was also a large proportion of unidentifiable sequence information within all three viral metagenomes that may encode genomic material of *Trichodesmium* cyanophages. The genome of *Trichodesmium erythraeum IMS101* contains CRISPR spacer sequences, which may indicate that this strain encountered phages in the past (Anderson et al., 2011). CRISPR spacer sequences are identified within the *T. erythraem* genome analyzed by MicrobesOnline, which is an analysis pipeline that identifies CRISPR elements missing from RefSeq annotations (Dehal et al., 2009). For these reasons, and based on previously published observations (Ohki, 1999; Hewson et al., 2004), it is likely that such a virus exists.

3.5 Conclusions

These data suggest that there is a diverse consortium of viruses associated with the collapse of *Trichodesmium* aggregations, but do not fully support the possibility that a *Trichodesmium* virus is responsible for aggregation collapse. There was a diverse group of viruses capable of infecting a number of different organisms present within collapsed *Trichodesmium* aggregations. Patterns of viral genotype representation reveal variations in host activity and virus infection strategies following *Trichodesmium* lysis. Viral genotypes that were more represented in the untreated library may represent lytic viruses of active opportunistic bacteria, while viruses that were better represented in the mitomycin C treated viral metagenomes may have been temperate viruses of abundant but slow growing hosts. While a virus of *Trichodesmium* could not be identified within these metaviromic datasets, *Trichodesmium*-like genes were present amongst sequence reads, and *Trichodesmium* genes were found in all libraries. Thus, based on this study there is inconclusive evidence that there is a virus of *Trichodesmium.* Higher sequencing coverage may be required to identify a virus of a dominant organism within a mixed microbial community. Future studies aimed at determining the cause of *Trichodesmium* aggregation collapse will require more directed approaches such as better isolation of the cyanobacterium away from associated bacteria, and deeper sequencing efforts.

3.6 References

Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410.

Anderson RE, Brazelton WJ, & Baross JA (2011) Using CRISPRs as a metagenomic tool to identify microbial hosts of a diffuse flow hydrothermal vent viral assemblage. *FEMS Microbiol Ecol* 77: 120–133.

Angly FE, Felts B, Breitbart M, et al. (2006) The marine viromes of four oceanic regions. *Plos Biol* 4:2121–2131.

Berman-Frank I, Bidle KD, Haramaty L & Falkowski PG (2004) The demise of the marine cyanobacterium, *Trichodesmium spp*., via an autocatalyzed cell death pathway. *Limnol Oceanogr* 49:997–1005.

Berman-Frank I, Rosenberg G, Levitan O, Haramaty L & Mari X (2007) Coupling between autocatalytic cell death and transparent exopolymeric particle production in the marine cyanobacterium *Trichodesmium*. *Environ Microbiol* 9:1415–1422.

Bhasvar J, Polson SW, Dhankar S & Wommack EK (2009) VIROME: A Resource for Analysis of Viral Metagenomes. *Abstracts of the General Meeting of the American Society for Microbiology* 109.

Capone DG, Zehr JP, Paerl HW, Bergman B & Carpenter EJ (1997) *Trichodesmium*, a globally significant marine cyanobacterium. *Science* 276:1221–1229.

Clokie MRJ, Millard AD & Mann NH (2010) T4 genes in the marine ecosystem: studies of the T4-like cyanophages and their role in marine ecology. *Virol J* 7:291–309.

Dehal PS, Joachimiak MP, Price MN, et al. (2009) MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res* 38:D396–D400.

Dyhrman ST, Chappell PD, Haley ST, Moffett JW, Orchard ED, Waterbury JB & Webb EA (2006) Phosphonate utilization by the globally important marine diazotroph *Trichodesmium*. *Nature* 439:68–71.

Gomez-Alvarez V, Teal TK & Schmidt TM (2009) Systematic artifacts in metagenomes from complex microbial communities. *ISME J* 3:1314-1317.

Hawser SP, Oneil JM, Roman MR & Codd GA (1992) Toxicity of blooms of the cyanobacterium *Trichodesmium* to zooplankton. *J Appl Phycol* 4:79–86.

Hewson I, Govil SR, Capone DG, Carpenter EJ & Fuhrman JA (2004) Evidence of *Trichodesmium* viral lysis and potential significance for biogeochemical cycling in the oligotrophic ocean. *Aquat Microb Ecol* 36:1–8.

Hewson I, Poretsky RS, Dyhrman ST, Zielinski B, White AE, Tripp HJ, Montoya JP & Zehr JP (2009) Microbial community gene expression within colonies of the diazotroph, *Trichodesmium*, from the Southwest Pacific Ocean. *ISME J* 3:1286–1300.

Hmelo LR (2010) Microbial interactions associated with biofilms attached to *Trichodesmium spp.* and detrital particles in the ocean. PhD Thesis, Massachusetts Institute of Technology - Woods Hole Oceanographic Institution, Joint Program in Oceanography/Applied Ocean Engineering, Woods Hole, MA

Ivars-Martinez E, Martin-Cuadrado AB, D'Auria G, Mira A, Ferriera S, Johnson J, Friedman R & Rodriguez-Valera F (2008) Comparative genomics of two ecotypes of the marine planktonic copiotroph Alteromonas macleodii suggests alternative lifestyles associated with different kinds of particulate organic matter. *ISME J* 2:1194–1212.

McDaniel L, Breitbart M, Mobberley J, Long A, Haynes M, Rohwer F & Paul JH (2008) Metagenomic Analysis of Lysogeny in Tampa Bay: Implications for Prophage Gene Expression. *PLoS One* 3:9

McDaniel LD, Houchin L, Williamson S & Paul JH (2001) Evaluation of seasonal changes in lysogeny in cyanobacteria in Tampa Bay. *Abstracts of the General Meeting of the American Society for Microbiology* 101:489.

Van Mooy BAS, Hmelo LR, Sofen LE, Campagna SR, May AL, Dyhrman ST, Heithoff A, Webb EA, Momper L & Mincer TJ (2012) Quorum sensing control of phosphorus acquisition in *Trichodesmium* consortia. *ISME J* 6:422–429.

Nagasaki K (2008) Dinoflagellates, diatoms, and their viruses. *J Microbiol* 46: 235–243.

Ng FFT, Wheeler E, Greig D, Waltzek TB, Gulland F & Breitbart M (2011) Metagenomic identification of a novel anellovirus in Pacific harbor seal (*Phoca vitulina richardsii*) lung samples and its detection in samples from multiple years. *J Gen Virol* 92:1318–1323.

O'Neil J & Roman M (1992) Grazers and associated organisms of *Trichodesmium. In Marine Pelagic Cyanobacteria: Trichodesmium and Other Diazotrophs. (*E.J. Carpenter, D.G. Capone, and J.G. Rueter, eds.) pp. 61-73 Bamberg: Kluwer Academic Publishers, Dordecht, The Netherlands.

Ohki K (1999) A possible role of temperate phage in the regulation of *Trichodesmium* biomass. *B I Oceanogr (Monaco)* 0:287–291.

Rho M, Tang H & Ye Y. (2010) FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res* 38:20

Rodier M & Le Borgne R (2010) Population and trophic dynamics of *Trichodesmium* thiebautii in the SE lagoon of New Caledonia. Comparison with *T. erythraeum* in the SW lagoon. *Mar Pollut Bull* 61:349–359.

Sheridan CC, Steinberg DK & Kling GW (2002) The microbial and metazoan community associated with colonies of *Trichodesmium spp,.*: a quantitative survey. *J Plankton Res* 24:913–922.

Shoaib M, Baconnais S, Mechold U, Le Cam E, Lipinski M & Ogryzko V (2008) Multiple displacement amplification for complex mixtures of DNA fragments. *BMC Genomics* 9:415

Simpson AA, Nandhagopal N, Van Etten JL & Rossmann MG (2003) Structural analyses of Phycodnaviridae and Iridoviridae. *Acta Crystallogr Sect D-Biol Crystallogr* 59:2053–2059.

Sullivan MB, Krastins B, Hughes JL, Kelly L, Chase M, Sarracino D & Chisholm SW (2009) The genome and structural proteome of an ocean siphovirus: a new window into the cyanobacterial "mobilome." *Environ Microbiol* 11:2935–2951.

Sun SL, Chen J, Li WZ, Altintas I, et al. (2011) Community cyberinfrastructure for Advanced Microbial Ecology Research and Analysis: the CAMERA resource. *Nucleic Acids R* 39:D546–D551.

Thingstad TF (2000) Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnol Oceanogr 45*, 1320–1328.

Thurber RV, Haynes M, Breitbart M, Wegley L & Rohwer F (2009) Laboratory procedures to generate viral metagenomes. *Nat Protoc* 4:470–483.

Vardi A, Van Mooy BAS, Fredricks HF, Popendorf KJ, Ossolinski JE, Haramaty,L & Bidle KD (2009) Viral Glycosphingolipids Induce Lytic Infection and Cell Death in Marine Phytoplankton. *Science* 326:861–865.

Weinbauer MG & Suttle CA (1999) Lysogeny and prophage induction in coastal and offshore bacterial communities. *Aquat Microb Ecol* 18:217–225.

Wilcox RM & Fuhrman JA (1994) Bacterial Viruses in Coastal Seawater- Lytic Rather than Lysogenic Production. *Marine Ecol-Prog Ser* 114:35–45.

Chapter 4. Comparison of water column virus populations over a 180+ year time period within the varved sediments of Fayetteville Green Lake

4.1 Abstract

Viruses of bacteria are abundant, diverse and dynamic members of aquatic ecosystems whose diversity is thought to be locally high but globally limited. The reasons behind this seemingly contradictory observation are likely due to two mechanisms that drive diversity of viral populations: mutation and dispersal. Understanding how viral populations change through time *in situ* will help elucidate the role of each process in the overall diversity of viral populations. This study examines the dynamics of viral populations over the course of 180+ years through examination of viruses within chronologically preserved in lake sediments. Fayetteville Green Lake (Green Lake) is a meromictic lake in Upstate New York that undergoes annual whiting events that lead to an undisturbed chronology of varved sediments within the deep basin of the lake. Dated sediment core samples were used to examine phages of cyanobacteria, a population of viruses that infect a host expected to occur exclusively in the surface waters. This study employed metagenomics and targeted biomarker sequencing to characterize the structure of populations of viruses within the sediments that originated in the water column. The metagenomic analyses indicated that cyanophage populations clustered by sediment age. Targeted sequencing of a conserved structural gene from sediments and the water column identified two clades of cyanobacterial myoviruses, one clade distributed within the sediments and water column, and another found only in the water column. Together these results suggest that cyanophage within Green Lake are temporally dynamic and the viral population contains genotypes that have persisted throughout the lake's history as well as a genotype that have recently dispersed into the lake. These data show that sediments serve as a repository of past populations of viruses within this lacustrine system, and may serve as a starting point for future investigations into the evolutionary ecology of *in situ* viral populations.

4.2 Introduction

It is becoming increasingly apparent that viruses impact ecosystem dynamics in diverse environments. Viruses mediate the conditions and constituents of the environments they inhabit through influencing host population dynamics, community level metabolic functioning and host community structure (Breitbart, 2012; Fuhrman, 1999; Hewson et al., 2013; Suttle, 2007; Weinbauer, 2004). Viruses play active roles in controlling diverse processes such as the collapse of phytoplankton blooms in the oligotrophic open ocean (Vardi et al., 2009, 2012) and targeted infection of specific members of the human microbiome (Barr et al., 2013). In order for viruses to carry out such critical and highly specific processes, however, they must establish themselves within a given environment and system and specifically adapt to the conditions and prey that they encounter. Given the cosmopolitan nature of viruses, this is a seemingly amazing achievement that is likely accomplished through rapid adaptation of the virus population to the conditions it encounters after dispersal, or by being on the winning end of a co-evolutionary arms race with their host population.

Viruses of bacteria are ideal candidates for the examination of mechanisms of evolution, adaptation and diversification because of their short generation times, potential for recombination with co-infecting phages, large numbers of progeny, and high probabilities of new genotypes on relatively short timescales (Krupovic et al., 2011). Elements of virus genomes are shared between habitats; nearly identical viral genes have been found in diverse environments, suggesting low global diversity (Angly et al., 2006; Breitbart et al., 2004). On a fine scale and over time, viral diversity is high. A temporal study of virus population structure using viral metagenomics revealed consistent coarse scale structure but dynamic changes in genotypic structure over daily scales, showing high levels of diversity at the nucleotide level (Rodriguez-Brito et al., 2010). Two studies that compared populations over the course of days to months to years found consistent diversity at each point of sampling, but varying population structure through time (Emerson et al., 2012, 2013).

How do viral populations change through time? Two mechanisms, viral dispersal and local mutation, are hypothesized to be major contributors. Separate studies of virus and host genomes within extreme environments provide evidence that both mechanisms are at play. A study that identified viral sequences based on host CRISPR spacer sequences within viral metagenomes from acid mine drainage found high levels of recombination within viral sequences, suggesting rapid adaptation to combat host immunity (Andersson and Banfield, 2008). A study examining *Sulfolobus* phage and host genomes revealed biogeographic patterns of prophages within the host genome, suggesting that interactions with local phage populations drove differences in host genomic diversity thus showing that mutation is important on a global scale (Held and Whitaker, 2009). In contrast, a group examining mechanisms of viral diversity in Yellowstone National Park hot springs over a two year period found that mutation alone could not explain the level of diversity and change within a given hot spring and that populations in seemingly isolated hot springs were not dispersal limited (Snyder et al., 2007). These contrasting results may be due to differences in temporal and spatial scale. In the laboratory, studies of phage-host relationships have shown that dispersal of phage between genetically distinct populations increases the adaptive capabilities of the phage relative to the host. This suggests that adaptation and phage dispersal may complement each other in maintaining phage diversity (Morgan et al., 2005).

The majority of *in situ* viral biomarker and metagenomic studies to date have spanned short timescales of days to at most several years. Studies that examine CRISPR spacer sequences within host genomes reveal a co-evolutionary history of virus and host populations over a presumably longer time period but inferences about the previously encountered phage population must be considered based only on small pieces of viral genomic information found within a CRISPR spacer (21 to 72 bp) (Barrangou et al., 2007; Sorek et al., 2008). Sediment cores are another way to look back in time at the evolutionary dynamics of planktonic populations because they harbor dormant or preserved organisms present over long periods of time. Revival of *Daphnia* from distinct

sediment layers showed evidence of co-evolution between a zooplankton host and its bacterial parasite (Decaestecker et al., 2007), and rapid adaptation in response to eutrophication (Hairston et al., 1999). These studies examined revived individuals stored in sediments up to half a century earlier. Sediments may also serve as a repository for changes in microbial diversity, such as changes in the viral population through time. Adsorption to sinking particles is considered an important mode of removal of viruses from aquatic habitats (Hewson and Fuhrman, 2003; Weinbauer, 2004), and several studies indicate that virus DNA and in some case intact viruses originating in the surface waters may persist deep within anoxic sediments for long periods of time as well. Coastal sediments near British Columbia contained lytic viruses capable of infecting the toxic marine alga, *H. akashiwo*, up to 40 cm below the sediment-water interface despite the occurrence of the host organism exclusively in surface waters (Lawrence et al., 2002). Phages capable of infecting the marine picocyanobacterium, *Synechococcus*, were recovered from deep in marine sediments in the Gulf of Mexico. Sedimentation rate estimates indicated that the sediment layers were 100 years old, suggesting that infectious cyanophage within the anoxic sediments had settled out of the water column a century earlier (Suttle, 2000). In the Black Sea, phage of surface water algal populations (*Emiliana huxleyi)* were observed to be present in 7000 year old sediment layers (Coolen, 2011). This study compared coarse scale population dynamics of marker genes for *E. huxleyi* and their viruses, stored within the sediments. The current study addresses viral community dynamics over 180+ years of history by examination of populations of phage that infect picocyanobacteria within chronologically distinct anoxic sediments of a freshwater lake.

The Study System

This investigation examines viruses that infect picocyanobacteria within Fayetteville Green Lake over a timescale of more than 180 years. Within Green Lake State Park in Fayetteville, NY, Green Lake and neighboring, biogeochemically similar and connected Round Lake, are an excellent

system to study *in situ* mechanisms of viral diversity and evolution because of their unusual biological and physical characteristics that make them a low-diversity, semi-isolated and contained system. Both lakes are small (c.a. 500 m diameter), permanently stratified, and sit in a small catchment basin (4.3 km^2) (Hilfinger et al., 2001). Green Lake is the better studied of the two lakes. Surface water within Green Lake is estimated to have residence times of 1.0 to 1.8 years (Brunskill and Ludlam, 1969). Permanent stratification of surface water from saline, anoxic bottom waters leads to nutrient poor surface waters that support a primary producer population dominated by the small, photosynthetic bacterium, *Synechococcus* (Schultze-Lam et al., 1992; Thompson and Ferris, 1990; Thompson et al., 1997). As the dominant planktonic primary producers, *Synechococcus* likely represent an important component of surface water biogeochemical cycling and food web dynamics within Green Lake.

Because Green Lake is small, deep and permanently stratified with anoxic bottom waters, it is an ideal location to find chronologically laminated (varved) sediments (Brunskill, 1969; Hilfinger et al., 2001; Ludlam, 1969, 1984). Sediment layers can be easily seen in Green Lake sediment cores due to annually occurring whiting events in which calcite ($CaCO_3$) precipitates from the water column (Ludlam, 1984). These annual whiting events are mediated by picocyanobacteria, and the onset of whiting events in Green Lake corresponds with a spring *Synechococcus* bloom within the lake (Thompson et al., 1997). A number of investigations have established a chronology for sediment cores by counting individual layers using x-radiography imaging and radioisotope dating (Hilfinger et al., 2001; Kirby et al., 2002; Ludlam, 1984). Characteristics of Green Lake sediments can be linked to historical land use and models of past climatological events and sediment cores from the basin of the lake date back 2500 years (Ludlam 1984, Hilfinger et al. 2001, Kirby et al. 2002).

Through examination of signatures of water column viruses within the Green Lake sediments originating from different periods in the lake's history, this study has two primary aims:

a) verify the presence of preserved cyanophage within the Green Lake sediments, and b) examine how cyanophage populations have changed through time within Green Lake to assess how mutation and dispersal influence the diversity and structure of cyanophage communities through time.

4.3 Methods

This study examines the signatures of viruses within Green Lake sediment cores using two complementary approaches: viral metagenomics and clone library sequencing of a cyanophage biomarker gene. In addition, it utilizes several different techniques for viral collection, purification and sequencing. All water column samples were collected between 2012 and 2014, and the sediment core samples were collected once on December 12, 2012.

Sediment Core Collection: Sediment cores were collected on December 12, 2012 from the basin of Fayetteville Green Lake. Two cores were taken in the southeastern portion of the deep basin (Cores B and C at 50m depth), and one core was taken from the northwestern part of the deep basin (Core A, 50m depth, see Appendix 2, Figure A2.2). Sediment cores were collected using a gravity corer and immediately covered with aluminum foil after collection to minimize light exposure. The air temperature on the day of sampling was 0^{o}C, and cores were kept between 0 and 4^{o}C during processing, except for a brief period of time for x-radiography, and during core sectioning. Cores were stored in the dark at 4^{o}C for 24 hours before being subjected to digital x-radiography at the Cornell University Veterinary Imaging facility. After imaging, they were returned to 4^{o}C until they were sectioned into 0.5 to 1cm thick cross-sections within 72 hours of collection. Briefly, core slices were pushed up through the core tube and sliced from the top using an ethanol-rinsed polyacetate sheet. The outer 0.5 cm perimeter of each section was removed to prevent contamination from the

corer tube, adjacent sections or the water column. Core slices were then transferred to sterile 50 ml conical tubes and stored at -80^{o}C until further processing.

X-radiographic images of sediment cores were analyzed using Carestream Direct digital imaging software (Rochester, NY) and a chronology was established based on manual varve counts of x-radiographic images. Sediment layer composition according to the x-ray images was matched to sediment sample depth measurements and notes taken on sediment composition during core slicing to pair sediment chronology with sediment core slices, following previously described methods (Hilfinger et al., 2001).

Detection of a cyanobacterial myovirus biomarker gene: Water samples for amplification and sequencing of the conserved cyanobacterial myovirus g20 gene were collected from the surface waters of Green Lake and Round Lake periodically during the summer of 2012 and 2013 as part of a related time series study of Green Lake and Round Lake (Brown Thesis Ch. 2, unpublished). Briefly, 1 L water from one site in Green Lake and one site in Round Lake was collected, transported to the laboratory and successively pre-filtered through a 10 µm polycarbonate (Isopore; Millipore) and 0.2 µm polyvinyldifluoride (Durapore; Millipore) filter and collected onto a 0.02 µm PVDF anotop filter (Whatman) using a peristaltic pump and swinnex filter holders for the 10 µm and 0.2 µm filters. Viral DNA was extracted from anotop filters using a Zymo Viral DNA extraction kit with a modified protocol as follows. Anotop filters were thawed and 800 µl of ZR viral buffer was introduced into the filter using a sterile 3 ml syringe. When the buffer filled the filter, anotops were plugged with a sterile flame-sealed micropipette tip, vortexed for 5 seconds and then incubated at room temperature for 10 minutes. After 10 minutes, the majority of the viral buffer was removed from the filter back into the syringe using backpressure, and the liquid in the syringe was transferred to a sterile 15mL conical tube. The filter was then pierced with a sterile sealed micropipette tip and the remaining liquid was pushed through the filter and into the same sterile

tube. The extraction then proceeded with the expunged viral buffer according to the manufacturer's protocol.

Bulk sediment DNA for phage biomarker sequencing was extracted from sediment layers using a Soil DNA Extraction kit according to the manufacturer's protocols (Zymo, Irvine, CA).

g20 Biomarker Sequencing: Extracted bulk DNA from sediment fractions and water column viral DNA from various time points in Round Lake and Green Lake were surveyed for the presence of the cyanobacterial myovirus portal protein, g20, using primers cps1.1 (5′-GTAGWATWTTYTAYATTGAYGTWGG-3′) and cps8.1 (5′-ARTAYTTDCCDAYRWAWGGWTC-3′) (Sullivan et al., 2008). PCR reactions contained 1× reaction buffer, 3.5 mM $MgCl_2$, 0.05U/µl Taq Polymerase (Promega), 1mM dNTPs (Promega PCR Nucleotide Mix), 2µl extracted viral DNA, 20 pg/µl bovine serum albumin (BSA), and 0.05 µM of each primer. Reactions began at 94°C for 3 minutes and then cycled through 35 rounds of 94°C melting temperature for 1 minutes, 35°C annealing temperature for 1 minute, 73°C extension for 1 minute. After 35 cycles the reactions were held at 73°C for 4 minutes for a final extension. PCR products were subjected to gel electrophoresis (1.5% agarose, 100V for 2 hours) and gels were stained with SYBR Gold to visualize reaction products. Sequencing of g20 sequences from water column samples was conducted in two ways. All sediment and select water column samples with PCR products of the expected size were chosen for cloning and sequencing. DNA was gel purified using the Zymo gel DNA extraction kit, cloned using the P-Gem T-easy vector system (Promega) and Sanger-sequenced at the Cornell University Biotecholonlogy Services Unit. Sequences were processed in CLC Genomics Workbench resulting in 109 clones from Round Lake and Green Lake from 2012, 42 clones from Green Lake in 2013 and 88 clones from Green Lake sediments (Table 2).

An additional g20 sequence library was created for Illumina sequencing to encompass as much g20 diversity as possible. PCR amplicons from several samples collected during the first three sampling time points (4/27, 5/2 and 5/10) of the 2012 time series from both Green Lake and

Round Lake were pooled and submitted for Illumina MiSeq. The resulting libraries were trimmed and denovo assembled in CLC Genomics Workbench (0.99 ID over 0.8 of the read). Nine complete g20 sequences were recovered. In addition, reads from this library were recruited against a clustered clone library of g20 sequences from the sediments and water column in CLC genomics workbench (0.95 ID over 0.7 of the read).

Phylogenetic Analysis of g20 sequences: g20 nucleic acid sequences were aligned in MEGA using the MUSCLE codon alignment algorithm (Edgar, 2004; Kumar et al., 2008). Two alignments were done using MUSCLE (Edgar, 2004), one with all g20 clone library sequences (used for diversity scores and Unifrac comparisons), and another with only representative sequence clustered at 99% identity using UCLUST. The alignments were hand-curated to remove all gaps and phylogenetic relationships were determined using the unweighted pair group method with arithmetic mean (UPGMA) algorithm available through MEGA. Trees were constructed using various tree-building algorithms available through MEGA. All trees revealed similar overall structure, but the UPGMA method yielded the most robust bootstrap values. Phylogenetic trees were visualized in itol (Letunic and Bork, 2007). Weighted and unweighted unifrac scores for the sequence alignment of the entire library were determined in mothur (Schloss et al., 2009). Translated amino-acid sequences from the water column of Green Lake and Round Lake and the sediment of Green Lake were combined with translated amino acid sequences of the cyanophage g20 gene collected in the Atlantic Ocean (Jameson et al., 2011), Lake Erie (Wilhelm et al., 2006), French Perialpine lakes (Dorigo et al., 2004; Zhong and Jacquet, 2013), Rice Paddy Soils (Jing et al., 2014), the Atlantic coast (Marston and Amrich, 2009; Marston and Sallee, 2003) and cultured representatives (Sullivan et al., 2008). Translated amino-acid sequences were deriplicated within each population dataset used. Sequences were aligned using MUSCLE (Edgar, 2004), curated and subjected to UPGMA phylogenetic analysis as described above.

DNA extraction and preparation for viral metagenomic sequencing: Water was collected from the basin of Green Lake at two time points: March 7, 2013 (20L) and May 7, 2014 (120L). Samples were pre-filtered through 10µm and 0.2µm filters using positive pressure filtration and concentrated using tangential flow filtration (TFF). Six sediment layers were selected for viral metagenome construction (Table 4.1): 1 near the surface sediments from a shallow core that had well defined varves at within the near-sediment water interface layers (SC11), two fractions of approximately the same chronology from two different locations in the lake (SB67 and SA64) and three chronologically distinct deep sediment fractions from the same sediment core (SB152, SB157 and SB184) from the deepest sediment core (See Appendix 2, Figure A2.3). Sediment samples were named in the following manner: the "S" stands for sediment, the second letter (A, B or C) represents the sediment core from which the sample was taken, and the number indicates the mean varve-count for the sample, representing the approximate age.

Three grams of sediment from each selected sediment fraction were combined with 3 ml of 10 mM sodium pyrophosphate ($NaPPO_4$, pH=7.5) and lightly sonicated for 5 min. The samples were spun down at 2500×g for 10 minutes, and then the supernatant was transferred to a new, sterile tube. This process was repeated on each sediment sample two more times, resulting in 9ml of supernatant per sample. The virus-containing supernatant was then centrifuged for 30 minutes at 2500×g to pellet any remaining suspended solids. Purified viruses from the water column and sediments were then filtered through a 0.2µm acrodisc filter (Pall company) and combined with polyethyleneglycol 8000 (PEG 8000) to a concentration of 10% PEG and were left to precipitate at 4^oC overnight. Samples were then pelleted at 11,000 rpm for 30 minutes, and resuspended in 1ml H_20. The May2014 sample was subjected to an additional round of purification using Cesium Chloride (CsCl) gradient centrifugation with density layers at 1.3g/ml, 1.5g/ml and 1.7g/ml. The layer around the interface between 1.5 and 1.7g/ml, expected to contain bacteriophage (Thurber et al., 2009), was extracted from the column for subsequent purification. 0.2 volumes of chloroform

were added to each, the samples were lightly mixed and left to incubate at room temperature for 10 minutes. Samples were then spun at 3000 rpm for 1 minute and the aqueous layer was transferred to a new tube. Next, samples were nuclease treated with 14 units of Turbo DNase, 20 units of RNase and Benzonase and incubated at 37°C for three hours after which EDTA was added to a final concentration of 20 mM to halt the nuclease activity. Viral DNA was then extracted using a Zymo viral DNA extraction kit according to the manufacturer's protocols.

Viral DNA was amplified using the Genomeplex® Whole Genome Amplification Kit (Sigma). The May2014 water column sample was handled according to the manufacturer's protocol. All other samples were subjected to a modified protocol to increase the final concentration of amplified DNA. Briefly, libraries were processed according to the manufacturer's protocol all the way through the amplification step. After amplification, a new master mix containing 30 µl amplification master mix diluted to 1× with the kit-provided water and 5 µl additional DNA polymerase were combined with 40 µl of the previous amplification product and thermal cycled according to the kit's amplification protocol for a second round of amplification. DNA was then purified using the Zymo DNA clean and concentrate kit and submitted for Illumina sequencing at the Cornell Biotechnology Resource Center's Genomics Core facility. The March 2013 water column and sediment samples SC11 and SB184 were subjected to 2×150 sequencing on Illumina MiSeq; sediment samples SB67, SA64, SB152 and SB157 were subjected to 2×250 sequencing on an Illumina MiSeq. The variability of sequencing method was dependent on the currently available sequencing technology. The May 2014 sample was subjected to 2×150 sequencing on an Illumina HiSeq and 2x250 sequencing on an Illumina MiSeq.

Comparison of viral metagenomes using clustering and read recruitment: All viral metagenomic libraries were trimmed and cleaned using prinseq (Schmieder and Edwards, 2011, -min_len 100 -min_qual_mean 15 -ns_max_p 5 -derep 14 -lc_method dust -lc_threshold 30), paired in mothur

(Schloss et al., 2009, make.contig), and both paired and non-paired sequences were combined and assembled in CLC genomics workbench using the de novo assembly option (QUIAGEN, 0.95 ID over 0.7 read length). Nucleic acid sequences of open reading frames (ORFs) greater than 200 bp were identified using "getorf" and translated into amino acid sequences using "transeq", both available through the "EMBOSS" package (Rice et al., 2000). All translated ORFs were annotated by BLASTp comparison to the CAMERA viral proteins database (e<0.001). For annotation, translated ORFs were compared to CAMERA Viral Proteins database (available at http://camera.crbs.ucsd.edu/reference-datasets/) with a minimum e-value of 10^{-3} using BLASTp (Altschul et al., 1997; Seshadri et al., 2007). The matched virus gene with the highest BLAST bitscore was used to annotate each ORF for which a match was found.

Several different databases were constructed by clustering open reading frames from assembled contigs of several different combinations of these seven libraries. SB67 was omitted from these analyses due to its small library size that prevented adequate information for comparison to other viral metagenomes. A water column (WC) database combining clustered ORFs from the two water column libraries (MayWC and MarchWC), a sediment database with clustered ORFs from sediment libraries SC11, SA64, SB152, SB157 and SB184 and a combined database containing the non-redundant water column and non-redundant sediment ORF libraries were all utilized to compare the representation of cyanophage across libraries. For each database, ORFs were clustered (0.95 nucleotide identity) using UCLUST (Edgar, 2010) and the longest representative from each cluster was kept for each non-redundant database.

Sediment read libraries SC11, SA64, SB152, SB157, SB184 and Green Lake water column libraries MayWC and MarchWC were compared to the different databases, described above, using read recruitment. To assign recruitment scores to each ORF within the different ORF database, 500,000 reads from each library were trimmed to 100 bp and then recruited to each database using CLC genomics workbench (0.95 identity over 0.7 of the sequence). Overall success of each database

for comparison between samples was determined by examination of the overall recruitment of reads, and recruitment of reads to cyanophage genomes for each read library against each database, with results indicated in Table 4.2. These methods are similar to methods used in a recent study examining the representation of specific gene groups and clusters in hypersaline Lake Tyrell (Emerson et al., 2013).

Representation and clustering of cyanophage genomes was carried out using the recruitment of reads to the combined ORF database. All recruitment scores for ORFs identified as most closely matching sequenced genomes of phages that infect *Synechococcus* and/or *Prochlorococcus* were summed for each cyanophage genome. Bray-Curtis distances were calculated in R using the vegan package (Oksanen et al., 2007, method= complete). A heat map depicting representation of cyanophage genomes within the seven compared viral metagenomes was drawn using the ggplot2 package (Wickham, 2009). Cyanophage diversity scores of Shannon diversity, the number of observed OTUs and Pielou's evenness were calculated using the vegan package (Oksanen et al., 2007) in R and are based on the amount of read recruitment from each normalized read library to ORFs mostly closely resembling genes from different cyanophage genomes as described above.

The relatedness of cyanophage signatures between viral metagenomes was determined by constructing a data matrix of recruitment scores for each ORF matching a cyanophage gene from the combined ORF database. Each ORF was assigned a recruitment score from each of the seven size- and length-normalized read libraries. Using this information, a Bray-Curtis distance matrix between samples was calculated using the vegan package in R (Oksanen et al., 2007). These scores were visualized via hierarchical clustering using the hclust base function in R (method=complete).

4.4 Results

Examination of cyanophage genes present within sediment and water column viral metagenomes

Five sediment and two water column viral metagenomes were compared to a database of assembled, non-redundant ORFs to determine the representation of different cyanophages through time in Green Lake. An additional water column library (IGR-g20) was used to further characterize the g20 clone library (Figure 4.2, Appendix 2, Figure A2.1, Table A2.1), and an additional sediment library (SB67) was too small to include in comparisons between libraries and was excluded from all analyses.

The seven compared samples include two collected from the water column of Green Lake, in March of 2013 (MarchWC) and May of 2014 (MayWC), and five libraries from sediment layers that differed in their distance from the surface as well as approximate age, estimated through counts of layers of annually deposited calcite (varves), imaged using x-radiography. Sediment layers were sectioned into 0.5 to 1 cm portions while varves were generally spaced <0.1 cm apart, but had regions of deposition of sandy or plant material that would increase the distance between varves. Distribution of interceding plant and sandy material was not consistent between sediment cores, causing the differences number of varves below the sediment-water interface between cores. This method of varve counting and the sediment core characteristics were consistent with a previous study that examined the chronology of layers within Green Lake sediment cores (Hilfinger et al., 2001). The approximate age of sediment sections used spanned 2 to 16 year periods depending on their depth below the surface and the varve distribution. The oldest layer (SB184) encompassed varves 176 to 192 (16 year span), and the shallowest sediment sample used (SC11) encompassed varves 9 to 13 (4 year span). Two sections were selected from differing cores that overlapped in varve counts (sample SB67 spanning varves 64 to 70, and sample SA64 which spanned varves 61 to 67) and two layers were selected from the same core that spanned temporally close periods of time (Sample SB152 [varves 150-153] and SB157 [varves 156-157], Table 4.1). Resulting read libraries

ranged in size from 162 thousand reads to 14 million reads and were sequenced using several different sequencing methods (Table 4.1). To account for differences in library size and sequencing methods, read length and library size was normalized between samples to compare phage representation across samples.

There was little shared sequence information between water column and sediment libraries. Between three and seven percent of reads from sediment libraries recruited to ORFs constructed from the MarchWC and MayWC viral metagenomes, and similarly four to six percent of reads from water column libraries recruited to ORFs from the sediment viral metagenomes. In contrast, 25 to 62% of reads from the normalized read libraries recruited to ORFs constructed from the same location (location being either the sediments or the water column). Representation of ORFs annotated as cyanophage was low in all viral metagenomes, but more so in the sediment libraries, where less than 0.75% of reads recruited to cyanophage-annotated ORFs. One to five percent of reads from the WC libraries recruited to cyanophage-annotated ORFs (Table 4.2). Due to differences in recruitment of reads from one location (such as the sediments) to the other location (such as the water column), the combined library containing non-redundant ORFs from all seven viral metagenomes was used as the reference database for which to examine the representation of cyanophage signatures between the different libraries.

ORFs resembling genes from 82 different sequenced cyanophage genomes were observed within the combined ORF database. These cyanophage genomes had varying levels of representation within the viral metagenomes, with clear clustering of cyanophage genomes based on abundance that appeared relatively consistent across samples (Figure 4.1A). When examining the relationship of cyanophage ORFs between viral metgenomes, viral metagenomes appeared to cluster based on sample location with the two deepest samples clustering together, the two mid-level samples being most closely related and the water column samples clustering together and

apart from the sediment libraries (Figure 4.1B). Estimates of cyanophage diversity, richness and evenness appeared relatively consistent between samples (Figure 4.1C).

Table 4.1. Illumina sequence library information. "Varve count" indicates the range of light-colored calcite layers from the surface of the sediment core, representing the approximate age of deposition, and depth refers to the depth below the sediment-water interface. *Libraries sequenced on a 2x150 run on an Illumina MiSeq . **Libraries sequenced on a 2x250 run on an Illumina MiSeq. ***Library subjected to additional purification step and sequenced on two separate machines, an Illumina HiSeq 2x150 run and an Illumina MiSeq 2x250 run. Additional information can be found in the methods section.

Sample Name	Sediment Core	Varve Count (depth)	Cleaned Reads	Contigs	Mean Contig Length	Mean Contig Coverage	# orfs >200bp
SC11*	C	9-13 (3 cm)	1480297	30685	468	3.74	51351
SB67**	B	64-70 (12.5 cm)	162835	5706	512	3.49	11240
SA64**	A	61-67 (13.75 cm)	2646068	105465	449	3.17	195748
SB152**	B	150-153 (26.5 cm)	619115	46998	502	2.25	103783
SB157**	B	156-157 (29.5 cm)	855359	89175	531	1.88	208041
SB184*	B	176-192 (33 cm)	1628319	41836	466	3.23	91925
MarchWC*	Surface WC		1510678	42534	441	2.34	85615
MayWC***	Surface WC		14789136	134799	626	19.13	306937

Table 4.2 Recruitment of normalized reads (500,000 reads, trimmed to 100bp per read) to the three databases constructed to examine representation of phage between libraries. ORFs within each of the databases were annotated using the results of a BLAST comparison of ORFs to the Camera Viral Proteins database.

Read Library:	MayWC	MarchWC	SC11	SA64	SB152	SB157	SB184
Database	% Read Recruitment to Database						
WC ORFs	62.39	24.67	4.88	6.91	5.65	4.84	3.07
Sediment ORFs	5.74	4.81	40.75	45.21	49.90	47.90	40.52
All ORFs	63.47	25.62	41.44	45.25	49.95	48.04	40.88
	% Read Recruitment to Cyanophage-annotated ORFs in Database						
WC ORFs	5.20	1.00	0.08	0.05	0.09	0.03	0.05
Sediment ORFs	0.13	0.06	0.47	0.23	0.73	0.45	0.67
All ORFs	5.18	1.01	0.49	0.23	0.71	0.47	0.70

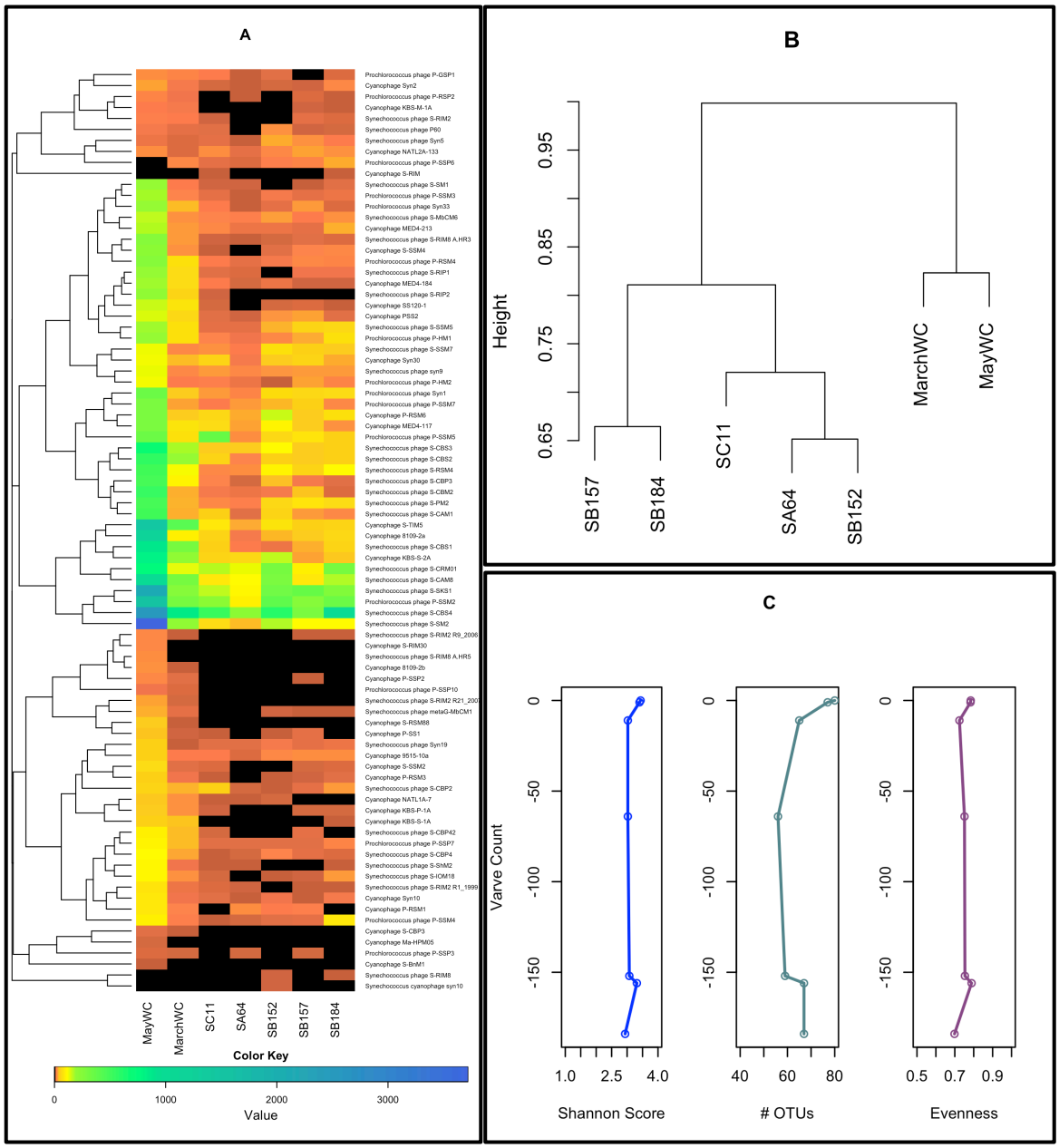


Figure 4.1 Representation of combined water column and sediment library open reading frames with best BLAST hit to cyanophage proteins pooled by cyanophage within the viral metagenomes SC11, SA64, SB157, SB152, SB184 and water column libraries MarchWC and MayWC. Representation of cyanophages (rows) are clustered based on Bray-Curtis clustering of cyanophage genome representation within each of the viral metagenomes, heatmap values are based on the number of reads that recruited to ORFs resembling genes from each cyanophage genome (A). Bray-Curtis dissimilarity scores of cyanophage-annotated ORF representation within each of the libraries represented by hierarchical clustering. Height indicates maximum distances between clusters (B). Diversity, richness and evenness estimates plotted against the number of varves below the sediment surface (approximate age) of the seven viral metagenomes. Shannon's diversity score, the number of observed OTU's and Pielou's Evenness were calculated based upon the representation of different cyanophage types within each of the viral metagenomes determined by read recruitment to cyanophage annotated ORFs, plotted based on the number of varves below the sediment surface (approximate age) (C).

Targeted sequencing of a conserved cyanomyovirus portal protein gene from Green Lake sediments and the water column of Green Lake and neighboring Round Lake

To identify cyanophage-related genes using a targeted approach, a conserved cyanophage portal protein found in cyanobacterial myoviruses (g20) was PCR amplified, cloned and sequenced from the water column of Green Lake, the water column of Round Lake, a biogeochemically similar lake that lies upstream of Green Lake, and from Green Lake sediment core layers. Extracted bulk DNA from many sediment layers were screened for clear amplification of the g20 marker gene using well established primers (Sullivan et al., 2008). g20 sequences were only present in four sediment layers: SA64, SA73, SA88 and SB152. These sediment layers fall between varves 61 and 153, and individual samples span four to six years in the history of Green Lake (Table 4.3). This sequencing effort resulted in 75 g20 sequences from the Green Lake surface waters, 69 sequences from Round Lake surface waters and 89 sequences from the Green Lake sediments (Table 4.3).

Green Lake and Round Lake surface water populations are significantly different from sediment populations (Unweighted Unifrac=0.8, $p<0.001$), and Green Lake sediments appear to have higher species richness than the water column (Chao richness: Green Lake/Round Lake water column combined=183 and combined sediments=251 at 100% similarity within OTUs). Phylogenetic analysis of nucleic acid aligned sequences reveals that cyanophage g20 sequences in this study generally fall into two large clades. Clade 1 contains sequences from the Green Lake sediments as well as from the water column, while Clade 2 contains sequences found exclusively in the water column. Within Clade 1, some branches encompass sequences only found in the sediments while others have representatives from both the water column and sediments (Figure 4.2).

Sequences from this study show similarity to cyanophage g20 sequences examined in previous studies of T4-like cyanophage diversity and biogeography. Sequences from Green Lake Clade 2 form their own branch within a larger group of uncultured freshwater g20 sequences.

Sequences from Green Lake Clade 1 are spread among sequences than span aquatic environments from marine to brackish to freshwater, in some cases sharing sub-clades with cultured cyanophage representatives (Figure 4.3).

g20 sequences present in two targeted water column viral metagenomes

No reads from the sediments and only one read from the MarchWC library recruited to the g20 clone library sequences (95% identity over 70% of the read), suggesting that viral metagenome coverage of these libraries was not high enough to capture this sequence. To get a better breadth of coverage of water column cyanophage sequences, reads from two additional Illumina sequenced libraries (IGR-g20 library and the MayWC library) were compared to the g20 clone library sequences.

Conserved cyanomyovirus portal protein sequences identical to those identified through the g20 clone library construction were obtained via assembly of the g20 IGR-g20 sequence reads. Additional g20 sequences not observed in the Green Lake libraries were also identified (Appendix 2 Figure A2.1, yellow). Presence of g20 sequence clusters based on read recruitment reflected the representation of water column viruses within the Green Lake/Round Lake water column in May 2012, but also indicates the presence of sequence clusters only observed in the sediment clone libraries. Fifty out of the 60 sequence clusters from the g20 clone library were represented within the IGR-g20 library (Figure 4.2, red color gradient).

Cyanophage portal protein sequences were also represented within the combined MayWC Illumina HiSeq and MiSeqlibraries, and reads from this library recruited to 35 out of 60 g20 sequence clusters. Patterns of abundance for the May2014 library differed from patterns seen in the IGR-g20 library, with less abundance of the Clade 2 sequences overall (Figure 4.2, magenta color gradient). Differences observed may reflect differences in community structure between years, but could also potentially reveal methodological differences in library construction. Regardless, determining the presence/absence of g20 clone library sequences within the water

column reveals that signatures of most g20 sequences found only in the sediments via clone library construction are also observed in the water column.

Table 4.3 Information about cyanophage g20 clone libraries and metagenomic libraries used to further characterize clone libraries. Additional information can be found in Appendix 2, Table A2.1.

Location Name	Core #	Depth (cm)	Varve Counts	# g20 clones	# Reads Recruited to clone library
SA64	A	13.75	61-67	21	
SA74	A	15.25	70-76	16	
SA88	A	17.75	84-91	30	
SB152	B	26.5	150-153	23	
Green Lake	Surface waters			75	
Round Lake	Surface waters			69	
Total Clones				234	
	Description				
MayWC	Surface Water viral metagenome				392
IGRL	Surface Water g20 Illumina library				13,175

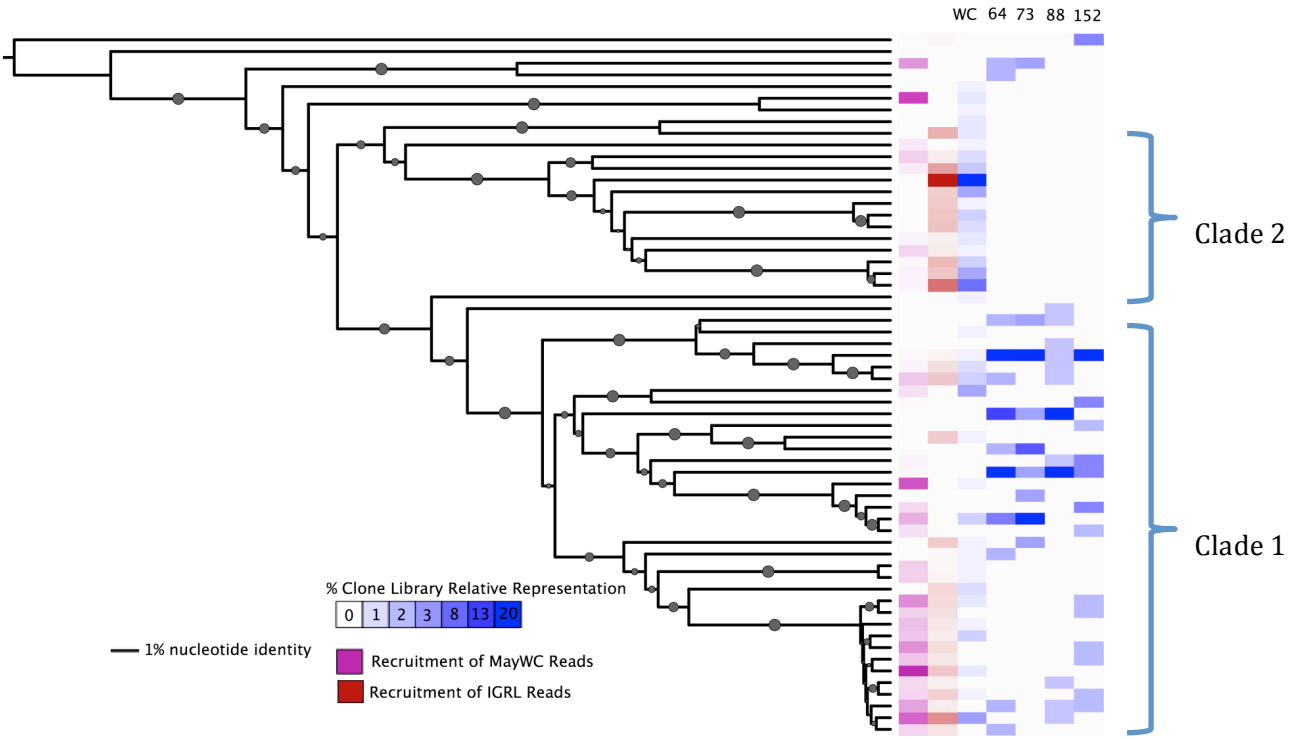
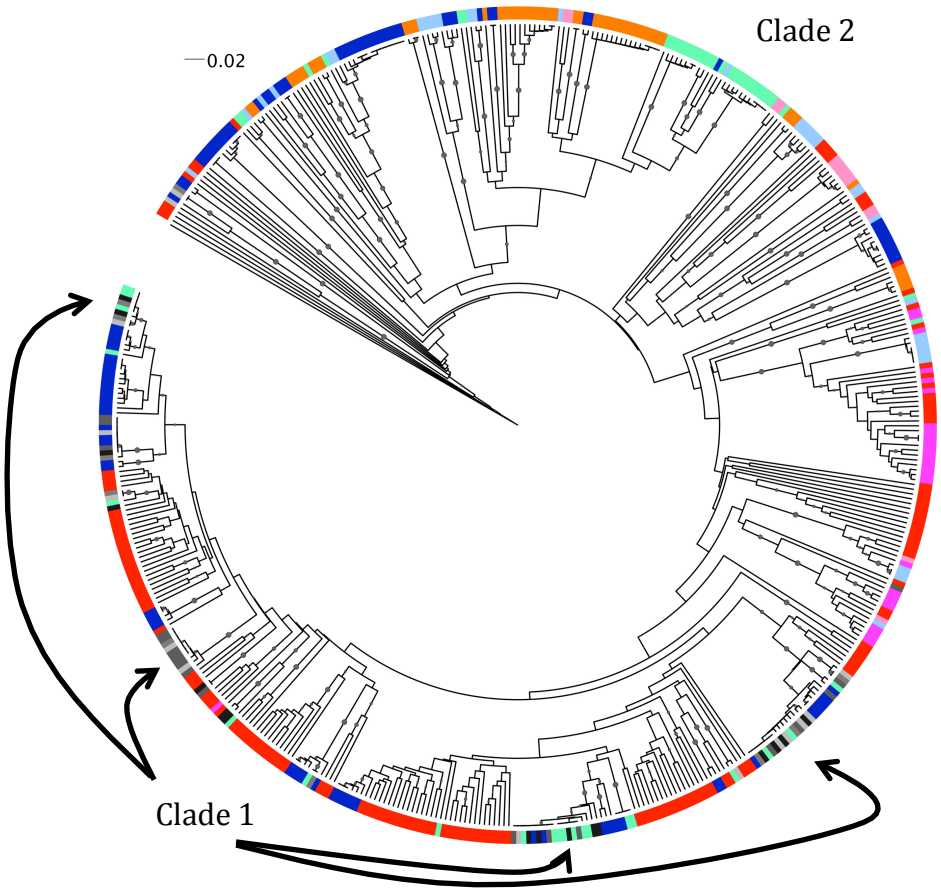


Figure 4.2 UPGMA phylogenetic tree of representative g20 nucleic acid sequences from the sediment and water column clone libraries clustered at 99% identity, aligned using MUSCLE and then subjected to UPGMA phylogenetic analysis with 500 iterations for bootstrap support. The blue heatmap on the right represents % relative representation of each cluster in the five clone libraries. For the water column (WC) n=153, for SA64 (64) n=21, for SA73 (73) n=16, for SA88 (88) n=30, for SB152 (152) n=23. The magenta and red color gradients represent the relative recruitment of reads from the MayWC and IGRL libraries to these clustered g20 sequences.



Location	Color	Source
Atlantic Ocean Transect		Jameson et al 2009
Rice Paddy Soils		Jing et al 2014
Coastal Atlantic Ocean		Maston and Sallee 2003, Marston and Amrich 2009
Cultured Representatives		Sullivan et al 2008
Green and Round Lake Surface Waters		This Study
Green Lake Sediment, 152 varves down		This Study
Green Lake Sediment, 64 varves down		This Study
Green Lake Sediment, 73 varves down		This Study
Green Lake Sediment, 88 varves down		This Study
Chesapeake Bay		Wang and Chen 2002
Lake Eerie		Wilhelm et al 2006
French Perialpine Lakes		Dorigo et al 2004, Zhong and Jaquet 2013
T4 Enteric Phage	No Color	

Figure 4.3 UPGMA phylogenetic tree of aligned g20 amino acid sequences from Green Lake and Round Lake water column, Green Lake sediments and from other geographically and biogeochemically diverse population datasets available through NCBI. Branches with >80% bootstrap support are noted by circles. Amino acid sequences from each population dataset were de-replicated using UCLUST at 100% identity.

4.5 Discussion

Cyanophage sequences are preserved in Green Lake sediments

Through the use of two complementary sequencing approaches, this study harnesses the chronology within sediment cores from varved, undisturbed sediments to examine current and past populations of viruses from Green Lake. Cyanophage are the primary focus of this study because their host population can only occur in the oxic surface waters of the lake and thus any signature of their phages seen in the sediments must have come from the surface waters at some point in the lake's history. Both shotgun sequencing of the bulk viral community and targeted sequencing of cyanophage marker genes reveals the presence of cyanophage genes within the deep sediments of Green Lake. All sediment viral metagenomic libraries had a number of ORFs with best BLAST match to cyanophage, and recruitment of reads to sequences with homology to cyanophage proteins further confirmed that all libraries possessed cyanophage sequences (Figure 4.1A). PCR amplification and sequencing of a conserved gene from cyanobacterial myoviruses, gene g20 encoding a portal protein, revealed the presence of this gene from sediment samples collected from portions of sediment cores that were 13 to 26 cm below the sediment-water interface (Figure 4.2).

Previous studies have used high throughput sequencing to explore potentially ancient viral DNA, such as in preserved coprolites (Appelt et al., 2014) and in stromatolites (Desnues et al., 2008), and previous studies have probed the dynamics of viral populations through time *in situ* (Emerson et al., 2012, 2013; Rodriguez-Brito et al., 2010). Viruses of surface water populations have been previously observed to be stored within sediments (Lawrence et al., 2002; Suttle, 2000), but this is the only the second study to date(the first being Coolen 2011 who examined *E. huxleyi* virus and host marker genes stored within Black Sea sediments), to examine molecular signatures of temporally stored virus populations within the sediment cores and the first of its kind to utilize viral metagenomics. As such, these data provide a unique view into the dynamics of sediment viral communities and their relationship to the deposition of viruses from the water column.

Trends with depth seen when comparing between sediment cyanophage populations

This study utilized the recruitment of length-normalized reads to a combined database of non-redundant ORFs from the individual assembly of each viral metagenomic library. Because of the chronology of the varved sediments, the expectation was that viruses in sediment layers that are in close proximity to each other would be more similar than layers that are further apart. Indeed, examination of the representation of cyanophage-annotated ORFs in the sediments and water column reveals that the two deepest sediment samples cluster together, the two mid-level samples are most similar to each other and the two water column viral metagenomes are most similar to each other (Figure 4.1B). Despite differences between samples, measures of diversity of represented cyanophage genomes in individual samples appeared consistent between libraries in the sediments and the water column (Figure 4.1C).

The low levels of recruitment of sediment reads to cyanophage ORFs from the water column viral metagenomes suggests that modern and past cyanophage populations within Green Lake are so different that defining past populations via comparison to the present population cannot reflect the diversity or relatedness of past populations (Table 4.2). Previous studies using viral metagenomic sequencing to examine viral populations over space and time have found similarly dynamic population structures; examination of protein clustering and genome abundance over the scale of months to years revealed changes in community structure despite consistent diversity of protein clusters (Emerson et al., 2013), and on an individual genomic level, viruses were persistent on a scale of days, but dynamic with transient presence on a timescale of years (Emerson et al., 2012). These results mirror observations from this study, as diversity appeared consistent when considering read recruitment to non-redundant ORFs from the sediments and the water column (Figure 4.1C).

Portal protein sequences found in Green Lake align with greater phylogeny of cyanomyovirus populations

In addition to shotgun sequencing of the bulk viral fraction within sediments and the water column, this study identified a conserved portal protein belonging to a subset of cyanophage, cyanobacterial myoviruses, within the sediments and water column of Green Lake. Previous studies of g20 sequence diversity show that cyanobacterial myovirus g20 genes are found in spatially and biogeochemically diverse environments (Sullivan et al., 2008). Looking at the Green Lake g20 sequence library in this larger context reveals interesting aspects of the Green Lake cyanobacterial myovirus populations. Clade 1 sequences from the water column and sediments fall in among sequences from diverse environments, including sequences obtained from cultured representatives. Clade 2 sequences form their own, albeit branched, sequence cluster within a larger clade of freshwater and terrestrial sequences. Without knowledge of their absence from the sediments, the unique phylogeny of Clade 2 Green Lake and Round Lake viruses would lead to the conclusion that this population is location specific within the Fayetteville Green Lakes. An alternative explanation is that this clade represents a regional clade of cyanobacterial myoviruses present in other freshwater lakes in the region, that recently took hold in Green Lake and Round Lake. Several sequences from other locations cluster with Clade 2 sequences: two sequences that cluster within Clade 2 are from a study of cyanophage populations in Lake Erie (Wilhelm et al., 2006) and several sequences from Lake Bourget, a French peri-alpine lake, are closely related (Zhong and Jacquet, 2013, Figure 4.3). The Clade 2 population of cyanomyoviruses is prevalent within the water column of Green Lake as evidenced by their high abundance within g20 clone libraries and a seasonal study monitoring their abundance (See Chapter 2). This study is one of very few studies to examine phage populations of freshwater lakes in the northeastern portion of North America. Further characterization of phage populations in the freshwater lakes of the region

and elsewhere would be required to determine the origin of this unique branch within the g20 cyanomyovirus phylogenetic tree, and perhaps the source of these recently introduced viruses.

g20 libraries reveal cyanomyovirus genotypes that are shared and others that are unique

Clades 1 and 2 show differing patterns of distribution between the sediment and water column. Clade 1 contains g20 sequences from both the water column and the sediments, while Clade 2 was found only in the water column (Figure 4.2).

Within Clade 1, there were several sequence clusters that contained only sequences from the sediments. Further comparison of all g20 clusters to two high-coverage targeted water column phage sequence libraries showed that some of the clusters with sediment-only cloned representatives could also be detected in the water column (Figure 4.2). As these two additional libraries were collected two years apart but during the same time of year, they account for variability among years, but do not encompass seasonal populations that are abundant at different points throughout the year. For example, one cluster of g20 sequences not observed in either of the high-coverage water column metagenomic libraries were only observed in samples collected in September (Appendix 2, Figure A2.1). It is possible that the sequences found exclusively in the sediments that were not detected in the high-coverage viral metagenomes are present in the water column during other seasons of the year, or at different depths within the mixolimnion water column. The presence of sediment g20 sequences in water column sequence libraries and similarities between water column clones and sediment clones shows that constituents of the cyanophage community have persisted over the span of 150 years represented in the g20 sequence library, while others may no longer exist in the lake.

The absence of deep-branching Clade 2 g20 sequences in the sediments, but abundance in the surface waters is a provocative result that suggests that this population was recently introduced into Green Lake and Round Lake. The shallowest sediment layer from which g20 sequences were amplified was layer SA64, a sediment layer that is approximately 61 to 67 years old according to

varve counts. Thus, if this hypothesis is correct, the Clade 2 population of cyanophage may have dispersed into Green Lake and Round Lake in the past 61 years. A previous study examining viruses of the marine algae, *E. huxleyi*, in the Black Sea found shifts in viral populations over the course of a 7000 year time period. In this record, some viral populations coincided with specific host populations whileother viral populations exhibited a transient existence throughout centuries, and the modern host population appeared in the sediment record with a different co-occurring viral population than that seen in modern times (Coolen, 2011). Clade 2 represents a potentially recent introduction of a unique population of water column phages that may indicate a shift in the relationship between modern viruses and cyanobacteria compared to the relationship between past populations.

Mutation and dispersal in Green Lake cyanophage populations

At least two alternative hypotheses could explain the absence of Clade 2 from the sediments. It is possible that the cyanomyoviruses represented by Clade 2 decay at a higher rate than Clade 1. If this is the case, then Clade 2 may have been historically present in the water column, but cannot be detected in the sediments. A previous study measuring the abundance of viruses of allochthonous and autochthonous origin in Green Lake and nearby Cayuga Lake found that different types of viruses had different decay rates in incubation experiments (Hewson et al., 2012) which were carried out in planktonic conditions. Generally, viral decay rates in sediments do not balance viral production rates, leading to high ratios of viruses to bacteria (Corinaldesi et al., 2010), and considering that both Clades come from the same viral family, it seems unlikely that they would decay at different rates. Alternatively, this water column population could have evolved to avoid deposition into the sediments. Calcite precipitation is an annually occurring phenomena that has been happening in Green Lake for at least 2500 years (Hilfinger et al., 2001). As viruses are known to adhere to sinking particles (Hewson and Fuhrman, 2003) and viruses may serve as a nucleation site for organomineralization (Pacton et al., 2014), it is possible that calcite precipitation

within Green Lake may lead to a significant loss of virus like particles from the water column. As such, the conditions within Green Lake may select for viruses that evade removal via particle adhesion or mineral nucleation. If this is the case, then Clade 2 cyanomyoviruses would not be recently introduced but rather a highly evolved and differentiated population.

Regardless of which hypothesis is correct, the results from this study suggest that both mutation and dispersal are influencing the diversity of cyanophage populations through time within Green Lake. With respect to the g20 clone library sequences, dispersal could potentially be occurring in both Clade 1 and Clade 2 populations. The evidence for Clade 2 is their presence in surface waters but absence from the sediments, which could easily lead to the conclusion that the Clade 2 populations are recently introduced. The widespread distribution of Clade 1 g20 sequences among sequences from diverse environments provides evidence that this group is globally dispersed. Persistence of viral types through time is evidenced by the presence of clusters of g20 sequences within 153-year-old sediments that also have representatives in younger sediment layers and the water column. Some sequences found in the deepest sediment layers are not observed in the clone libraries of shallower sediments, or the water column, but are closely related to sequences from younger sediments and the water column (Figure 4.2). The sequences found exclusively in the deep sediments could potentially represent sequences from ancestral viruses no longer present in the lakes that are directly related to members of the modern cyanophage population. This suggests that local mutation is also contributing to cyanophage diversity.

Viral metagenomes reveal high diversity and dynamic populations of cyanophage. High diversity within the viral metagenomes led to low coverage of the sediment virus community, which prevented direct comparisons of fine scale genetic differences between populations in the current study. Regardless, the clustering of populations from temporally similar deep sediment samples suggests that there is consistency in population structure through time, but change over longer timescales of decades. The large differences between surface water and sediment

populations suggests that the current population of cyanobacteria is different than historical populations. Fine scale evolutionary analyses are not within the scope of this study, but investigation of evolutionary rates within sub-clades of g20 sequences and investigations into more cyanophage marker genes within the sediments would help elucidate the evolutionary trajectory and dynamics driving the diversity of cyanophage populations within Green Lake.

Annotation-based characterization of viral metagenomes

Annotation-based analyses of viral metagenomes is a method that has been used since the beginning of viral metagenomics, but has limitations. Common issues associated with this method of analysis include database biases and the prevalence of unidentifiable sequences within viral metagenomes (Hurwitz et al., 2014). These two issues taken together lead to the majority of sequences being unidentifiable, and possible mis-annotation of sequences that are identified. Whether annotation or other methods of sequence characterization are used is totally dependent upon the questions asked. As the aim of this study was to probe populations within the sediments that originated in the water column, identification of sequences with homology to cyanobacterial viruses was critical. To increase the amount of information that could be gained from the short read sequence data, this study used read-recruitment to quantify the representation of specific viral groups in the sediments after cyanophage-annotated sequences were identified (Emerson et al., 2013).

4.6 Conclusion

This study investigates viral populations of a meromictic lake over 180 years of the lake's history using viral metagenomics and targeted clone libraries to probe current and historical populations of cyanophage. This two pronged approach lead to complementary insights into drivers of diversity within Green Lake cyanophage populations. Differences between the current surface water cyanophage populations compared to those found in the sediments suggest that the community structure of cyanophages within Green Lake is constantly changing. The presence of an entire Clade of viruses in the water column not observed in the sediments leads to greater discussion of drivers of viral diversity. The presence of such a population could lead to contrasting conclusions; either a recent dispersal event occurred or this population is highly evolved to exist within the unique conditions of Green Lake. Meanwhile, other subsets of the cyanophage population have persisted through time suggesting that localized populations are also present. These data show that signatures from water column populations of viruses from the lake's history are preserved in the sediments. In systems such as Green Lake where lake sediments are chronologically preserved, these populations may provide insights into mechanisms that influence how viral and host populations evolve and change through time.

4.7 References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

Andersson, A.F., and Banfield, J.F. (2008). Virus population dynamics and acquired virus resistance in natural microbial communities. Science *320*, 1047–1050.

Angly, F.E., Felts, B., Breitbart, M., Salamon, P., Edwards, R.A., Carlson, C., Chan, A.M., Haynes, M., Kelley, S., and Liu, H. (2006). The marine viromes of four oceanic regions. PLoS Biol. *4*, e368.

Appelt, S., Fancello, L., Le Bailly, M., Raoult, D., Drancourt, M., and Desnues, C. (2014). Viruses in a 14th-century coprolite. Appl. Environ. Microbiol. *80*, 2648–2655.

Barr, J.J., Auro, R., Furlan, M., Whiteson, K.L., Erb, M.L., Pogliano, J., Stotland, A., Wolkowicz, R., Cutting, A.S., and Doran, K.S. (2013). Bacteriophage adhering to mucus provide a non–host-derived immunity. Proc. Natl. Acad. Sci. *110*, 10771–10776.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709–1712.

Breitbart, M. (2012). Marine viruses: truth or dare. Mar. Sci. *4*.

Breitbart, M., Miyake, J.H., and Rohwer, F. (2004). Global distribution of nearly identical phage-encoded DNA sequences. FEMS Microbiol. Lett. *236*, 249–256.

Brunskill, G. (1969). Fayetteville Green Lake, New York. II. Precipitation and sedimentation of calcite in a meromictic lake with laminated sediments. Limnol Ocean. *14*, 830–847.

Brunskill, G., and Ludlam, S. (1969). Fayetteville Green Lake, New York. I. Physical and chemical limnology. Limnol Ocean. *14*, 817–829.

Coolen, M.J. (2011). 7000 years of Emiliania huxleyi viruses in the Black Sea. Science *333*, 451–452.

Corinaldesi, C., Dell'Anno, A., Magagnini, M., and Danovaro, R. (2010). Viral decay and viral production rates in continental-shelf and deep-sea sediments of the Mediterranean Sea. FEMS Microbiol. Ecol. *72*, 208–218.

Decaestecker, E., Gaba, S., Raeymaekers, J.A., Stoks, R., Van Kerckhoven, L., Ebert, D., and De Meester, L. (2007). Host–parasite "Red Queen"dynamics archived in pond sediment. Nature *450*, 870–873.

Desnues, C., Rodriguez-Brito, B., Rayhawk, S., Kelley, S., Tran, T., Haynes, M., Liu, H., Furlan, M., Wegley, L., and Chau, B. (2008). Biodiversity and biogeography of phages in modern stromatolites and thrombolites. Nature *452*, 340–343.

Dorigo, U., Jacquet, S., and Humbert, J.-F. (2004). Cyanophage diversity, inferred from g20 gene analyses, in the largest natural lake in France, Lake Bourget. Appl. Environ. Microbiol. *70*, 1017–1022.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics *26*, 2460–2461.

Emerson, J.B., Thomas, B.C., Andrade, K., Allen, E.E., Heidelberg, K.B., and Banfield, J.F. (2012). Dynamic viral populations in hypersaline systems as revealed by metagenomic assembly. Appl. Environ. Microbiol. *78*, 6309–6320.

Emerson, J.B., Thomas, B.C., Andrade, K., Heidelberg, K.B., and Banfield, J.F. (2013). New approaches indicate constant viral diversity despite shifts in assemblage structure in an Australian hypersaline lake. Appl. Environ. Microbiol. *79*, 6755–6764.

Fuhrman, J.A. (1999). Marine viruses and their biogeochemical and ecological effects. Nature *399*, 541–548.

Hairston, N.G., Lampert, W., Cáceres, C.E., Holtmeier, C.L., Weider, L.J., Gaedke, U., Fischer, J.M., Fox, J.A., and Post, D.M. (1999). Lake ecosystems: rapid evolution revealed by dormant eggs. Nature *401*, 446–446.

Held, N.L., and Whitaker, R.J. (2009). Viral biogeography revealed by signatures in Sulfolobus islandicus genomes. Environ. Microbiol. *11*, 457–466.

Hewson, I., and Fuhrman, J. (2003). Viriobenthos production and virioplankton sorptive scavenging by suspended sediment particles in coastal and pelagic waters. Microb. Ecol. *46*, 337–347.

Hewson, I., Barbosa, J.G., Brown, J.M., Donelan, R.P., Eaglesham, J.B., Eggleston, E.M., and LaBarre, B.A. (2012). Temporal dynamics and decay of putatively allochthonous and autochthonous viral genotypes in contrasting freshwater lakes. Appl. Environ. Microbiol. *78*, 6583–6591.

Hewson, I., Ng, G., Li, W., LaBarre, B.A., Aguirre, I., Barbosa, J.G., Breitbart, M., Greco, A.W., Kearns, C.M., and Looi, A. (2013). Metagenomic identification, seasonal dynamics, and potential transmission mechanisms of a Daphnia-associated single-stranded DNA virus in two temperate lakes. Limnol Ocean. *58*, 1605–1620.

Hilfinger IV, M.F., Mullins, H.T., Burnett, A., and Kirby, M.E. (2001). A 2500 year sediment record from Fayetteville Green Lake, New York: evidence for anthropogenic impacts and historic isotope shift. J. Paleolimnol. *26*, 293–305.

Hurwitz, B.L., Westveld, A.H., Brum, J.R., and Sullivan, M.B. (2014). Modeling ecological drivers in marine viral communities using comparative metagenomics and network analyses. Proc. Natl. Acad. Sci. *111*, 10714–10719.

Jameson, E., Mann, N.H., Joint, I., Sambles, C., and Mühling, M. (2011). The diversity of cyanomyovirus populations along a North–South Atlantic Ocean transect. ISME J. *5*, 1713–1721.

Jing, R., Liu, J., Yu, Z., Liu, X., and Wang, G. (2014). Phylogenetic Distribution of the Capsid Assembly Protein Gene (g20) of Cyanophages in Paddy Floodwaters in Northeast China. PloS One *9*, e88634.

Kirby, M., Patterson, W., Mullins, H., and Burnett, A. (2002). Post-Younger Dryas climate interval linked to circumpolar vortex variability: isotopic evidence from Fayetteville Green Lake, New York. Clim. Dyn. *19*, 321–330.

Krupovic, M., Prangishvili, D., Hendrix, R.W., and Bamford, D.H. (2011). Genomics of bacterial and archaeal viruses: dynamics within the prokaryotic virosphere. Microbiol. Mol. Biol. Rev. *75*, 610–635.

Kumar, S., Nei, M., Dudley, J., and Tamura, K. (2008). MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief. Bioinform. *9*, 299–306.

Lawrence, J.E., Chan, A.M., and Suttle, C.A. (2002). Viruses causing lysis of the toxic bloom-forming alga Heterosigma akashiwo (Raphidophyceae) are widespread in coastal sediments of British Columbia, Canada. Limnol. Oceanogr. *47*, 545–550.

Letunic, I., and Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics *23*, 127–128.

Ludlam, S.D. (1969). Fayetteville Green Lake III. The laminated sediments. Limnol Ocean. *14*, 848–857.

Ludlam, S.D. (1984). Fayetteville Green Lake, New York, USA: VII. Varve chronology and sediment focusing. Chem. Geol. *44*, 85–100.

Marston, M.F., and Amrich, C.G. (2009). Recombination and microdiversity in coastal marine cyanophages. Environ. Microbiol. *11*, 2893–2903.

Marston, M.F., and Sallee, J.L. (2003). Genetic diversity and temporal variation in the cyanophage community infecting marine Synechococcus species in Rhode Island's coastal waters. Appl. Environ. Microbiol. *69*, 4639–4647.

Morgan, A.D., Gandon, S., and Buckling, A. (2005). The effect of migration on local adaptation in a coevolving host–parasite system. Nature *437*, 253–256.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M.H.H., Oksanen, M.J., and Suggests, M. (2007). The vegan package. Community Ecol. Package.

Pacton, M., Wacey, D., Corinaldesi, C., Tangherlini, M., Kilburn, M.R., Gorin, G.E., Danovaro, R., and Vasconcelos, C. (2014). Viruses as new agents of organomineralization in the geological record. Nat Commun *5*.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. Trends Genet. *16*, 276–277.

Rodriguez-Brito, B., Li, L., Wegley, L., Furlan, M., Angly, F., Breitbart, M., Buchanan, J., Desnues, C., Dinsdale, E., Edwards, R., et al. (2010). Viral and microbial community dynamics in four aquatic environments. ISME J *4*, 739–751.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., and Robinson, C.J. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. *75*, 7537–7541.

Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. Bioinformatics *27*, 863–864.

Schultze-Lam, S., Harauz, G., and Beveridge, T. (1992). Participation of a cyanobacterial S layer in fine-grain mineral formation. J. Bacteriol. *174*, 7971–7981.

Seshadri, R., Kravitz, S.A., Smarr, L., Gilna, P., and Frazier, M. (2007). CAMERA: a community resource for metagenomics. PLoS Biol. *5*, e75.

Snyder, J.C., Wiedenheft, B., Lavin, M., Roberto, F.F., Spuhler, J., Ortmann, A.C., Douglas, T., and Young, M. (2007). Virus movement maintains local virus population diversity. Proc. Natl. Acad. Sci. *104*, 19102–19107.

Sorek, R., Kunin, V., and Hugenholtz, P. (2008). CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat. Rev. Microbiol. *6*, 181–186.

Sullivan, M.B., Coleman, M.L., Quinlivan, V., Rosenkrantz, J.E., DeFrancesco, A.S., Tan, G., Fu, R., Lee, J.A., Waterbury, J.B., and Bielawski, J.P. (2008). Portal protein diversity and phage ecology. Environ. Microbiol. *10*, 2810–2823.

Suttle, C.A. (2000). In The Ecology of Cyanobacteria, (Dordrecht, Netherlands: Kluwer Academic), pp. 563–589.

Suttle, C.A. (2007). Marine viruses—major players in the global ecosystem. Nat. Rev. Microbiol. *5*, 801–812.

Thompson, J., and Ferris, F. (1990). Cyanobacterial precipitation of gypsum, calcite, and magnesite from natural alkaline lake water. Geology *18*, 995–998.

Thompson, J.B., Schultze-Lam, S., Beveridge, T.J., and Des Marais, D.J. (1997). Whiting events: biogenic origin due to the photosynthetic activity of cyanobacterial picoplankton. Limnol. Oceanogr. *42*, 133–141.

Thurber, R.V., Haynes, M., Breitbart, M., Wegley, L., and Rohwer, F. (2009). Laboratory procedures to generate viral metagenomes. Nat. Protoc. *4*, 470–483.

Vardi, A., Van Mooy, B.A., Fredricks, H.F., Popendorf, K.J., Ossolinski, J.E., Haramaty, L., and Bidle, K.D. (2009). Viral glycosphingolipids induce lytic infection and cell death in marine phytoplankton. Science *326*, 861–865.

Vardi, A., Haramaty, L., Van Mooy, B.A., Fredricks, H.F., Kimmance, S.A., Larsen, A., and Bidle, K.D. (2012). Host–virus dynamics and subcellular controls of cell fate in a natural coccolithophore population. Proc. Natl. Acad. Sci. *109*, 19327–19332.

Weinbauer, M.G. (2004). Ecology of prokaryotic viruses. FEMS Microbiol. Rev. *28*, 127–181.

Wickham, H. (2009). ggplot2: elegant graphics for data analysis (Springer).

Wilhelm, S.W., Carberry, M.J., Eldridge, M.L., Poorvin, L., Saxton, M.A., and Doblin, M.A. (2006). Marine and freshwater cyanophages in a Laurentian Great Lake: evidence from infectivity assays and molecular analyses of g20 genes. Appl. Environ. Microbiol. *72*, 4957–4963.

Zhong, X., and Jacquet, S. (2013). Prevalence of viral photosynthetic and capsid protein genes from cyanophages in two large and deep perialpine lakes. Appl. Environ. Microbiol. *79*, 7169–7178.

Chapter 5. Conclusion

5.1 Summary of Results

Phages played a defining role in understanding fundamental concepts in molecular biology (e.g. Benzer, 1955; Brenner et al., 1961; Hershey and Chase, 1952), yet the high abundance of viruses in aquatic ecosystems was only realized in the past 30 years (Bergh et al., 1989; Proctor and Fuhrman, 1990). Since then, we have learned that viruses are abundant and important contributors to ecosystem dynamics (Fuhrman, 1999; Suttle, 2007). Although the field has made great strides in understanding role of viruses in ecosystems functions, the complexity and diversity of viruses have made them particularly difficult to characterize (Edwards and Rohwer, 2005). This dissertation contributes to the overall understanding of viral ecology through investigation into bacteriophage responses to varying scales of change. Chapter 2 addressed the replicability of the seasonal abundance of three cyanophages in similar lakes and compared virus seasonal abundance to the abundance of the host population. This study found consistent trends in cyanophage abundance between lakes, but annual variability and a weak relationship between the abundance of virus and host populations. Chapter 3 examined how the viral community responds to a sharp environmental change in the form of a cyanobacterial aggregation collapse and observed both lytic and lysogenic phages of bacteria known to be associated with *Trichodesmium* colonies. Chapter 4 examined viral populations over extended periods of time through identification of phage DNA within sediments and found interesting patterns of representation of genetic signatures of cyanophages in sediments spanning a 180+ year time period. The approach for each chapter was to investigate phage populations *in situ* and as such, there was little constraint on the many environmental variables that influence the dynamics of the investigated populations. In addition to the direct contributions of this research to our understanding of viral ecology, the broad scope of theses studies provides additional hypotheses and considerations for future investigations into viral community ecology.

In chapter 2 I found only weak evidence that supported the hypothesis that the changes viral and host populations are related and similar between lakes. The changes in abundance through time of three cyanophage types were similar between lakes but not between years. In addition, there was not a strong relationship between the three cyanophage types and members of the cyanobacterial population. This suggests that other variables are controlling the cyanobacterial population. I discuss grazers as one possible explanation for differing host dynamics, but it is also possible that other unmeasured members of the phage population control host population dynamics, or that host abundance is too low to support a predator population of phages. This study is the first in the lacustrine system to assess the dynamics of viruses and hosts in near replicate lakes. Considering there were differences in the seasonal dynamics of virus and host populations even within these two connected, biogeochemically similar lakes, future investigations into viruses and hosts within lacustrine systems must consider virus-host relationships within a well-defined lake environment rather than comparing across lakes of differing trophic levels. There are likely no universal trends within aquatic environments, but rather virus-host relationships that are dependent on unique environmental conditions. Complicating things further, the dynamics of viral populations may also be dependent upon outside factors such as rainfall patterns and anthropogenic manipulation (Van Der Gucht et al., 2001; Williamson et al., 2014).

In chapter 3, I found a diverse consortium of viruses associated with *Trichodesmium* aggregation collapse. Some members of the community that were more represented in uninduced incubations appeared to be lytic viruses of *Trichodesium*-associated bacteria likely responding to changes in the host's metabolism due to the influx of organic matter associated with *Trichodesmium* lysis. Other viruses that were more represented in the mitC-induced libraries resembled viruses capable of lysogeny, suggesting that this lifestyle is prevalent within *Trichodesmium*-associated bacterial genomes. The abundance of reads resembling genes for integrase proteins in both the

induced and uninduced libraries suggested that *Trichodesmium* lysis likely caused the induction of prophages present in the genomes of bacteria associated with *Trichodesmium* colonies.

The results in chapter 4 shows strong evidence of recent dispersal events and suggests that mutation is also playing a role in influencing the structure of cyanophage populations within Green Lake. I found that signatures from cyanophage populations stored within the sediments cluster with similarly aged sediment layers and are different than current populations of cyanophage in the surface waters. In addition, through examination of a conserved cyanophage biomarker gene, I identified two viral populations with differing distributions between the sediments and the water column, suggesting that subsets of the cyanophage population have differing evolutionary histories, with one population seen consistently throughout the lake's history, and the other either recently introduced, prone to decay or evolved to avoid deposition in the sediments. This study is the first to use viral metagenomics to examine historical populations of viruses within sediments, and shows that chronologically undisturbed sediments may be used to examine changes in virus populations through time.

5.2 Possible future directions

Observation of water column virus signatures within the GL sediments opens the door to a number of investigations into the evolution of viral communities and virus-host relationships. One possible avenue to pursue is to use the sediments of GL to assess the co-evolutionary relationships between viruses and bacteria. For example, information about past virus populations could be used in conjunction with bacterial genomics or metagenomics to examine the *in situ* dynamics of CRISPR sequence integration into microbial genomes. The sediments provide a temporal record of past planktonic viral populations and CRISPRs provide a record of past viruses encountered by a bacterium or archaeon (Barrangou et al., 2007; Horvath et al., 2008). Through comparison of historical populations of viruses from the sediments to CRISPR spacer sequences within genomes of

members of the planktonic microbial populations, this system could be used to determine the *in situ* rate of integration of CRISPR spacer sequences into CRISPR regions of microbial genomes. The study of these age-specific sediment viral metagenomes coupled with CRISPR spacer sequences of modern microbial populations could also address hypotheses formed in previous CRISPR studies, such as the idea that historical blooms of virulent viruses lead to conserved spacers within the older regions of CRISPR genomic elements (Weinberger et al., 2012).

Another possible line of investigation is characterization of the viral communities associated with various water column depths within GL using viral metagenomics. A thorough understanding of viral diversity in all parts of the water column would provide a wider picture of the full extent of viral preservation within lake sediments, as well as comparison of viral diversity across gradients. The geochemistry of GL and RL leads to a permanently stratified water column with the surface layer (mixolimnion) and bottom layer (monimolimnion) having very different geochemical characteristics. Thus GL and RL have a number of unique environments that host different microbial communities, including a thick mat of green- and purple-sulfur bacteria and purple non-sulfur bacteria at the chemocline that borders the aerobic to anaerobic transition between the mixolimnion and monimolimnion(Thompson et al., 1990). Comparison of viral communities between different horizons within GL and RL would also provide interesting biogeographic insights into viral community structure. While surface water residence time within the two lakes is around two years, monimolimnion water is predicted to stay at least 10-fold longer within each basin, with no obvious source of water connecting the monimolimnion of GL to that of RL (Hilfinger IV et al., 2001; Torgersen et al., 1981). Although there appeared to be sharing of cyanophage types between GL and RL in chapters 2 and 4, there may not be the same correspondence within the deeper, anoxic waters. Thus for such a study, I would hypothesize that the surface waters would have a larger group of shared viral genotypes, and a more similar community structure, while the deep waters would have more lake-specific separation in

community structure due to longer residence times and fewer opportunities for dispersal of monimolimnion bacteriophage between lakes.

There is also the question of *how* the viruses from the water column reach the sediments within GL. The whole reason I was able identify virus signatures from various periods of time in the lake's history is because of annual whiting events, the precipitation of calcite out of solution by planktonic *Synechococcus* populations occurring in the GL mixolimnion (Thompson and Ferris, 1990; Thompson et al., 1997), which lead to seasonally deposited calcite layers within lake sediments. The presence of water column viruses in the sediments proves that they are indeed transported from the mixolimnion into the deep waters and sediments. Do they adhere to calcite particles or organic matter? Decreases in viral abundance in the surface waters during the summer season follow the cyanobacterial spring bloom and correspond with the documented onset and duration of whiting events within GL (Thompson et al., 1997). As viruses are observed to adhere to suspended particles (Hewson and Fuhrman, 2003), it follows that the increase in calcite particles beginning in the spring and continuing through the summer leads to enhanced attachment and removal of viruses from the planktonic portion of the water column. In addition, it is possible that viruses in the water column actually contribute to the precipitation of calcite through providing a nucleation site for calcite mineralization. Calcite mineralization by GL *Synechococcus* is carried out in the cyanobacterium's outer S-layer (Schultze-Lam et al., 1992), the same region of the cell where phages would adhere to their host at the onset of infection. Perhaps some cyanophage serve as a site for calcite precipitation when in close proximity to their hosts. A recent study of viruses within microbial mats suggests that viruses can provide nucleation sites for organomineralization. This study found hexagonally shaped carbonate nanospheres within the mats, which the authors hypothesized were fossilized viral particles (Pacton et al., 2014). Green Lake would be a promising location to further investigate the possibility of viral preservation in the fossil record. Mineralized

viruses could potentially be found within calciferous sediment layers or within the carbonate bioherms that have formed on the perimeter of the lake.

Finally, there is currently a lack of information about how viral communities respond over time to rapid change within the environment. Chapter 3 of this dissertation begins to investigate this question through examination of viral communities immediately after *Trichodesmium* collapse and found high representation of viruses infecting gammaproteobacteria such as *Alteromonas*, known to respond to organic inputs in the marine environment (Ivars-Martinez et al., 2008). The next step would be to investigate how the virus community "recovers" from such shifting conditions. Does the prevalence of planktonic phages capable of lysogeny change after the initial pulse of organic matter, or do they permanently shift to a lytic lifestyle? Experiments that sample the viral and bacterial communities periodically after an input of organic matter, such as after *Trichodesmium* aggregation collapse, would help elucidate how viruses impact the cycling of nutrients in aquatic systems. Because phage infection leads to host cell lysis and thus a release of host organic matter into the environment, I would hypothesize that viral dynamics would follow a cyclical pattern after the onset of environmental change, with members of the viral community expanding in abundance due to host responses to specific pools of organic matter. In this manner, bacteriophages may mediate the succession of microbial species in response to organic inputs.

5.3 A synthesizing discussion of the results

Chapters 2 and 4 both examined the dynamics of cyanophage in Fayetteville Green Lake and to a lesser extent, Round Lake: Chapter 2 examined a small subset of the cyanophage population over a two year period, and Chapter 4 examined portions of the entire community for over 180 years. Both chapters observed changing populations of cyanophage. The variability in cyanophage abundance during two adjacent years within GL and RL offers an explanation of one environmental factor that could be influencing the variability in historical cyanophage populations in the GL

sediments. As total monthly rainfall was different between years and described some of the variability within cyanophage population dynamics in Chapter 2, differences in community structure through time may be driven by changes in climatological patterns. Indeed, based on examination of the entire annotated viral communities within the metagenomic data from chapter 4 the sediment sections most distantly related to other sediment layers occurred between 60 and 70 years ago which corresponds with a period of higher temperatures in the region, corroborated by isotope data from Hilfinger 2001 (Figure 5.1). Another piece of support for this theory is the ability to amplify g20 genes only from a handful of sediment layers, including sediment layers centered around the highlighted period of time between 64 and 88 years ago.

Studies observe that warmer temperatures shift the composition and activity of primary producer populations within lakes (Petchey et al., 1999; Sommer and Lengfellner, 2008), and

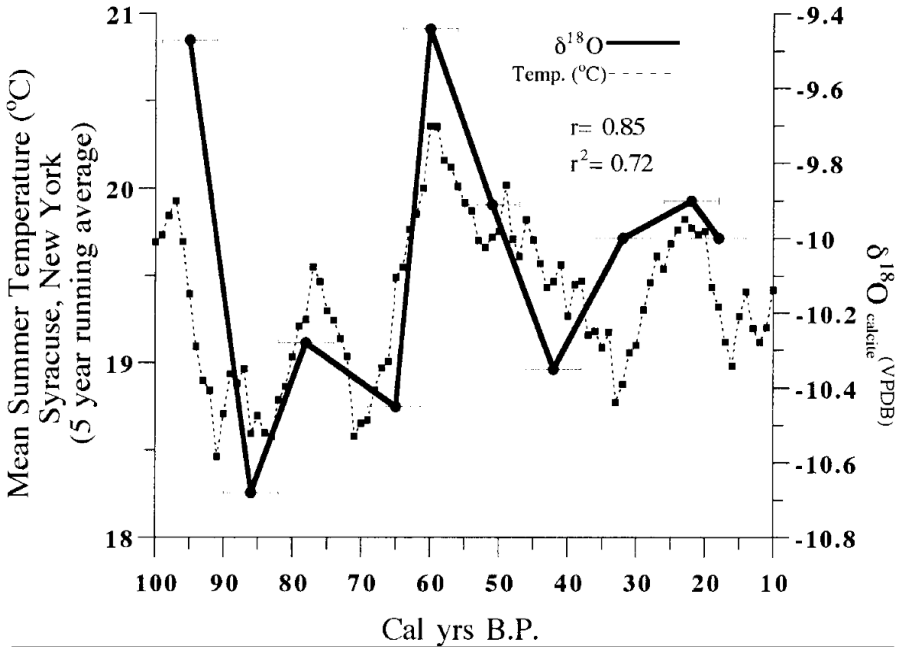


Figure 5.1 From Hilfinger 2001: "$\delta^{18}O$ calcite values from Green Lake core compared with mean summer air temperatures measured at nearby Syracuse, New York. Note strong positive correlation (r = 0.85) between $\delta^{18}O$ calcite values and mean summer air temperatures".

Note that this study took place 13 years before the current study, so the cal years would be shifted back by 10 years, and the period of warm lake temperatures and high $\delta^{18}O$ values would correspond to ~70 Cal yrs B.P. instead of 60.

temperature related changes are also observed in bacterial communities (Adams et al., 2010). Chapter 3 showed that sudden changes in the environment likely lead to a response in the viral community that are associated with shifts in the activity of the bacterial community. In Green Lake, changes in the primary producer dynamics in response to warmer temperatures may have led to the differing viral population structure within the 70 year-old sediment sections. Would shifts in infection strategies of these phage groups cause long-term differences in the viral community structure after this environmental change, or does the population eventually return to its previous state? Based on the findings in Chapter 4, community structure appears to be unidirectional as no sediment cyanophage populations resembled modern populations of viruses (Chapter 4, Figure 4.1, Table 4.2). The two oldest sediment cyanophage populations clustered most closely together, water column viruses clustered together and the two mid-level samples, which happen to also be the samples from which a g20 gene was able to be amplified, clustered together (Chapter 4 Figure 4.1B). Diversity estimates in the sediments show similar levels of diversity in all samples examined, despite differences in community structure. Together these observations suggest that on the short term of several years viral diversity may shift due to environmental change, but overall diversity is resilient through time within this oligotrophic environment.

5.4 Final thoughts

The field of viral ecology has expanded quickly since the discovery of abundant viruses within the environment. Despite this expansion, the field is small enough that viral ecologists are able to compare results across environments. Due to the complexity of viral communities and cosmopolitan nature of virus types (Breitbart and Rohwer, 2005), all observations made regardless of the environment enhance our overall understanding of viral ecology. One challenge will be drawing universal theories about what influences the composition of viral communities from observations from diverse environments. The findings in this thesis contribute to our overall understanding of viral communities across environments. For example, through finding variability

in the relationship between virus and host populations in biogeochemically similar environments, this work shows that generalizations cannot be made about the expected relationship between individual virus and host populations within complex environments.

Results from this dissertation indicate that viral communities responding to change show representation of prophage populations as well as high representation of lytic bacteriophage infecting metabolically responsive hosts. Thus when changes in the environment occur, both lytic and induced lysogenic populations may respond. This in turn could influence the composition of successional populations of bacterial hosts responding to environmental change. As viruses influence metabolic processes of microbial communities (Fuhrman, 1999; Suttle, 2007; Thompson et al., 2011), understanding how viruses respond to rapid environmental change will aid in understanding of overall ecosystem responses. In addition, these findings will not only inform understanding of virus populations in aquatic systems, but also in other systems experiencing sudden nutrient inputs. For example, these findings could provide insight into how viruses will affect the microbial population in a human gut after the mammalian host has consumed a large amount of labile carbon, such as fructose or ethanol.

Finally, through examination of viral populations over an extended period of time, this work observed recently dispersed populations of phage that have established themselves as one of the dominant populations in GL surface waters. Understanding how dispersed phages will impact endemic populations of viruses is a concept that may be applied across environments. One applied example of this is phage therapy, or the use of specific phages to control pathogentic populations of bacteria within a mammalian host (Letarov et al., 2010). Introduction of a new phage targeting an abundant bacterial population to a community could have unintended consequences and may lead to shifts in the over all structure of the viral community. Analysis of cyanophage populations within the GL sediments show that viral populations are continuously changing, and that migration may be an important contributor to overall viral diversity, even in seemingly isolated environments.

The future of the field is ripe for discovery. With new technologies for sequencing longer nucleic acid reads (Clarke et al., 2009; Eid et al., 2009), single cell and single virion sorting and sequencing (Allen et al., 2011; Lasken, 2012; Stepanauskas, 2012) and innovative methods for the concentration, identification and characterization of viruses (Hurwitz et al., 2014; Willner and Hugenholtz, 2013), many unsolved questions will soon be addressed. Discoveries of new and interesting viruses, such as DNA-RNA hybrid viruses (Diemer, 2014; Hewson et al., 2013; Roux et al., 2013), giant viruses of microbial eukaryotes (Claverie et al., 2006), and viruses that infect larger viruses (La Scola et al., 2008) offer new and intriguing insights into the evolution and origins of viruses. Investigations of viruses present in new and diverse environments offer exciting developments in the nature of viral communities. For example, bacteriophages associated with the mammalian microbiome indicate the possibility of three-tiered co-evolutionary relationships between phage, bacteria and mammalian hosts (Barr et al., 2013; Pride et al., 2011; Willner et al., 2011). Through providing insights into how viral communities change over several scales of inquiry, the work presented in this dissertation adds to the growing body of knowledge of viruses in the environment and offers unique insights that will greatly contribute to the field of viral ecology.

5.4 References

Adams, H.E., Crump, B.C., and Kling, G.W. (2010). Temperature controls on aquatic bacterial production and community dynamics in arctic lakes and streams. Environ. Microbiol. *12*, 1319–1333.

Allen, L.Z., Ishoey, T., Novotny, M.A., McLean, J.S., Lasken, R.S., and Williamson, S.J. (2011). Single virus genomics: a new tool for virus discovery. PloS One *6*, e17722.

Barr, J.J., Auro, R., Furlan, M., Whiteson, K.L., Erb, M.L., Pogliano, J., Stotland, A., Wolkowicz, R., Cutting, A.S., and Doran, K.S. (2013). Bacteriophage adhering to mucus provide a non–host-derived immunity. Proc. Natl. Acad. Sci. *110*, 10771–10776.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709–1712.

Benzer, S. (1955). Fine structure of a genetic region in bacteriophage. Proc. Natl. Acad. Sci. U. S. A. *41*, 344.

Bergh, Ø., BØrsheim, K.Y., Bratbak, G., and Heldal, M. (1989). High abundance of viruses found in aquatic environments. Nature *340*, 467–468.

Breitbart, M., and Rohwer, F. (2005). Here a virus, there a virus, everywhere the same virus? Trends Microbiol. *13*, 278–284.

Brenner, S., Jacob, F., and Meselson, M. (1961). An unstable intermediate carrying information from genes to ribosomes for protein synthesis. Nature *190*, 576–581.

Clarke, J., Wu, H.-C., Jayasinghe, L., Patel, A., Reid, S., and Bayley, H. (2009). Continuous base identification for single-molecule nanopore DNA sequencing. Nat. Nanotechnol. *4*, 265–270.

Claverie, J.-M., Ogata, H., Audic, S., Abergel, C., Suhre, K., and Fournier, P.-E. (2006). Mimivirus and the emerging concept of "giant" virus. Virus Res. *117*, 133–144.

Van Der Gucht, K., Sabbe, K., De Meester, L., Vloemans, N., Zwart, G., Gillis, M., and Vyverman, W. (2001). Contrasting bacterioplankton community composition and seasonal dynamics in two neighbouring hypertrophic freshwater lakes. Environ. Microbiol. *3*, 680–690.

Diemer, G.S. (2014). The Boiling Springs Lake Metavirome: Charting the Viral Sequence-Space of an Extreme Environment Microbial Ecosystem.

Edwards, R.A., and Rohwer, F. (2005). Viral metagenomics. Nat. Rev. Microbiol. *3*, 504–510.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., and Bettman, B. (2009). Real-time DNA sequencing from single polymerase molecules. Science *323*, 133–138.

Fuhrman, J.A. (1999). Marine viruses and their biogeochemical and ecological effects. Nature *399*, 541–548.

Hershey, A.D., and Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. J. Gen. Physiol. *36*, 39–56.

Hewson, I., and Fuhrman, J. (2003). Viriobenthos production and virioplankton sorptive scavenging by suspended sediment particles in coastal and pelagic waters. Microb. Ecol. *46*, 337–347.

Hewson, I., Ng, G., Li, W., LaBarre, B.A., Aguirre, I., Barbosa, J.G., Breitbart, M., Greco, A.W., Kearns, C.M., and Looi, A. (2013). Metagenomic identification, seasonal dynamics, and potential transmission mechanisms of a Daphnia-associated single-stranded DNA virus in two temperate lakes. Limnol Ocean. *58*, 1605–1620.

Hilfinger IV, M.F., Mullins, H.T., Burnett, A., and Kirby, M.E. (2001). A 2500 year sediment record from Fayetteville Green Lake, New York: evidence for anthropogenic impacts and historic isotope shift. J. Paleolimnol. *26*, 293–305.

Horvath, P., Romero, D.A., Coûté-Monvoisin, A.-C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., and Barrangou, R. (2008). Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. J. Bacteriol. *190*, 1401–1412.

Hurwitz, B.L., Westveld, A.H., Brum, J.R., and Sullivan, M.B. (2014). Modeling ecological drivers in marine viral communities using comparative metagenomics and network analyses. Proc. Natl. Acad. Sci. *111*, 10714–10719.

Ivars-Martinez, E., Martin-Cuadrado, A.-B., D'Auria, G., Mira, A., Ferriera, S., Johnson, J., Friedman, R., and Rodriguez-Valera, F. (2008). Comparative genomics of two ecotypes of the marine planktonic copiotroph Alteromonas macleodii suggests alternative lifestyles associated with different kinds of particulate organic matter. ISME J. *2*, 1194–1212.

Lasken, R.S. (2012). Genomic sequencing of uncultured microorganisms from single cells. Nat. Rev. Microbiol. *10*, 631–640.

Letarov, A., Golomidova, A., and Tarasyan, K. (2010). Ecological basis for rational phage therapy. Acta Naturae *2*, 60.

Pacton, M., Wacey, D., Corinaldesi, C., Tangherlini, M., Kilburn, M.R., Gorin, G.E., Danovaro, R., and Vasconcelos, C. (2014). Viruses as new agents of organomineralization in the geological record. Nat Commun *5*.

Petchey, O.L., McPhearson, P.T., Casey, T.M., and Morin, P.J. (1999). Environmental warming alters food-web structure and ecosystem function. Nature *402*, 69–72.

Pride, D.T., Salzman, J., Haynes, M., Rohwer, F., Davis-Long, C., White, R.A., Loomer, P., Armitage, G.C., and Relman, D.A. (2011). Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome. ISME J. *6*, 915–926.

Proctor, L.M., and Fuhrman, J.A. (1990). Viral mortality of marine bacteria and cyanobacteria.

Roux, S., Enault, F., Bronner, G., Vaulot, D., Forterre, P., and Krupovic, M. (2013). Chimeric viruses blur the borders between the major groups of eukaryotic single-stranded DNA viruses. Nat. Commun. *4*.

Schultze-Lam, S., Harauz, G., and Beveridge, T. (1992). Participation of a cyanobacterial S layer in fine-grain mineral formation. J. Bacteriol. *174*, 7971–7981.

La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., Fournous, G., Merchat, M., Suzan-Monti, M., Forterre, P., and Koonin, E. (2008). The virophage as a unique parasite of the giant mimivirus. Nature *455*, 100–104.

Sommer, U., and Lengfellner, K. (2008). Climate change and the timing, magnitude, and composition of the phytoplankton spring bloom. Glob. Change Biol. *14*, 1199–1208.

Stepanauskas, R. (2012). Single cell genomics: an individual look at microbes. Curr. Opin. Microbiol. *15*, 613–620.

Suttle, C.A. (2007). Marine viruses—major players in the global ecosystem. Nat. Rev. Microbiol. *5*, 801–812.

Thompson, J., and Ferris, F. (1990). Cyanobacterial precipitation of gypsum, calcite, and magnesite from natural alkaline lake water. Geology *18*, 995–998.

Thompson, J.B., Ferris, F., and Smith, D.A. (1990). Geomicrobiology and sedimentology of the mixolimnion and chemocline in Fayetteville Green Lake, New York. Palaios 52–75.

Thompson, J.B., Schultze-Lam, S., Beveridge, T.J., and Des Marais, D.J. (1997). Whiting events: biogenic origin due to the photosynthetic activity of cyanobacterial picoplankton. Limnol. Oceanogr. *42*, 133–141.

Thompson, L.R., Zeng, Q., Kelly, L., Huang, K.H., Singer, A.U., Stubbe, J., and Chisholm, S.W. (2011). Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. Proc. Natl. Acad. Sci. *108*, E757–E764.

Torgersen, T., Hammond, D., Clarke, W., and Peng, T. (1981). Fayetteville, Green Lake, New York: 3H-3He water mass ages and secondary chemical structure. Limnol Ocean. *26*, 110–122.

Weinberger, A.D., Sun, C.L., Pluciński, M.M., Denef, V.J., Thomas, B.C., Horvath, P., Barrangou, R., Gilmore, M.S., Getz, W.M., and Banfield, J.F. (2012). Persisting viral sequences shape microbial CRISPR-based immunity. PLoS Comput. Biol. *8*, e1002475.

Williamson, K.E., Harris, J.V., Green, J.C., Rahman, F., and Chambers, R.M. (2014). Stormwater runoff drives viral community composition changes in inland freshwaters. Front. Microbiol. *5*.

Willner, D., and Hugenholtz, P. (2013). From deep sequencing to viral tagging: recent advances in viral metagenomics. Bioessays *35*, 436–442.

Willner, D., Furlan, M., Schmieder, R., Grasis, J.A., Pride, D.T., Relman, D.A., Angly, F.E., McDole, T., Mariella, R.P., and Rohwer, F. (2011). Metagenomic detection of phage-encoded platelet-binding factors in the human oral cavity. Proc. Natl. Acad. Sci. *108*, 4547–4553.

Appendix 1. Supplemental Material for Chapter 2

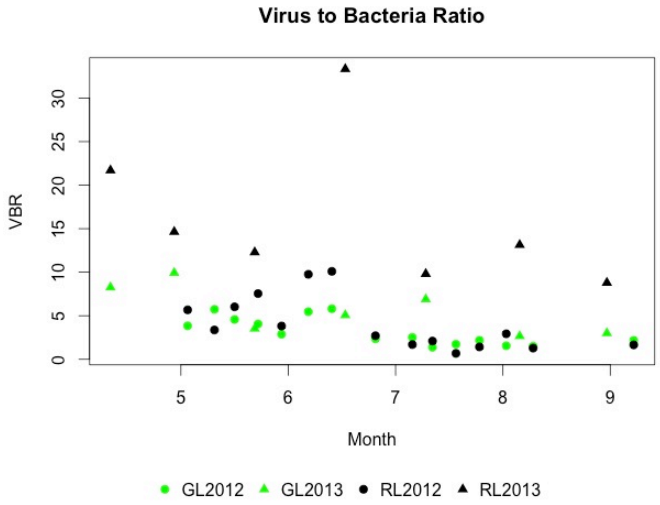


Figure A1.1. Virus to bacteria ratio (VLP ml^{-1}/bacteria ml^{-1}) in GL and RL in 2012 and 2013.

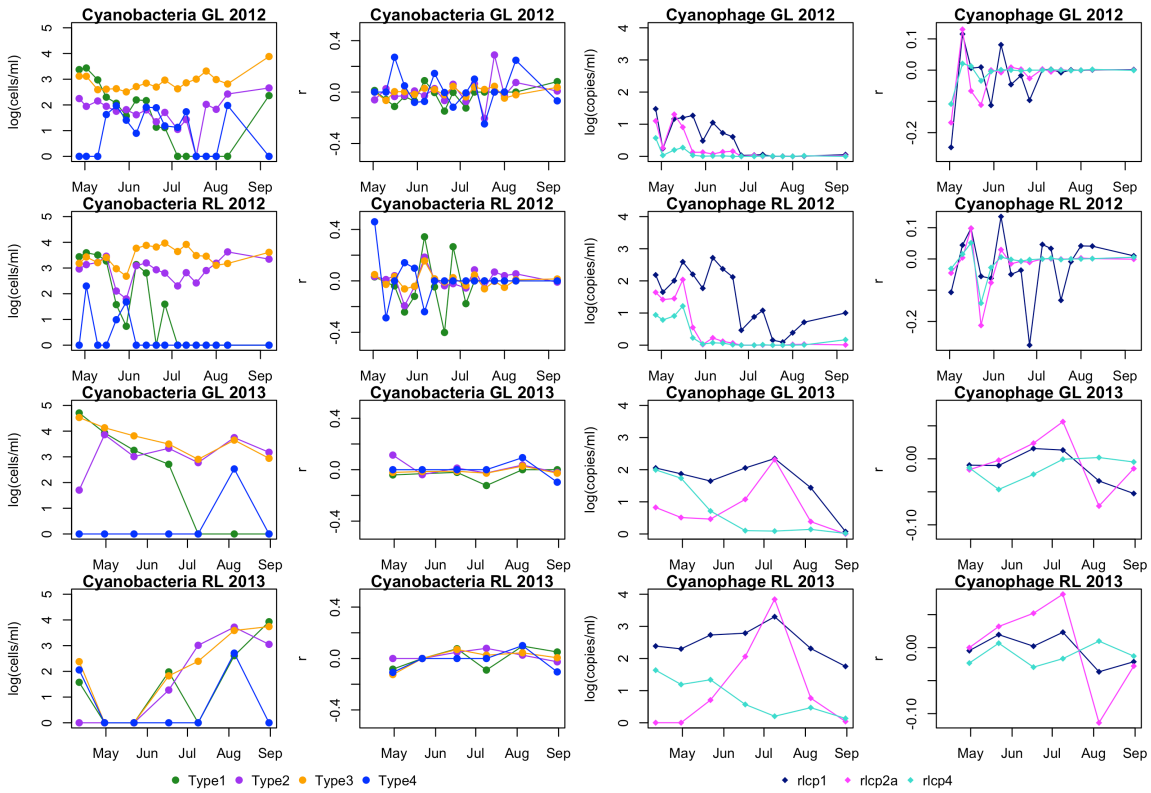


Figure A1.2 Cyanobacterial abundance (left column), cyanobacterial growth rate (left center column), cyanophage abundance (right center column) and cyanophage growth rate (right column). Units for growth rate (r) is calculated as log(organisms ml^{-1} day^{-1}) for cyanobacteria and log(copies ml^{-1} day^{-1}) for cyanophage types.

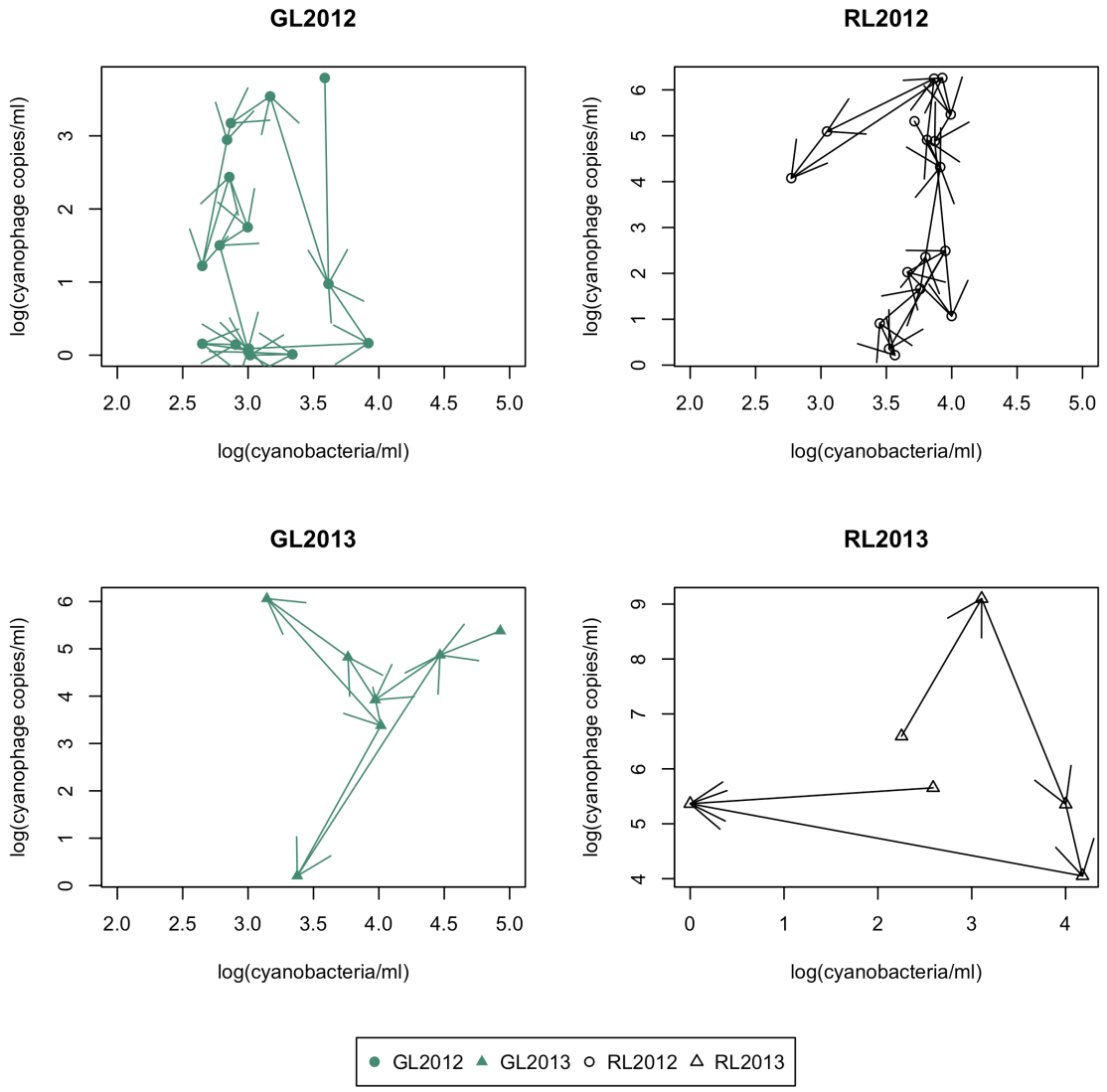


Figure A1.3 Cyanobacterial abundance plotted against cyanophage abundance in GL and RL in 2012 and 2013. Arrows indicate chronology of each set of values.

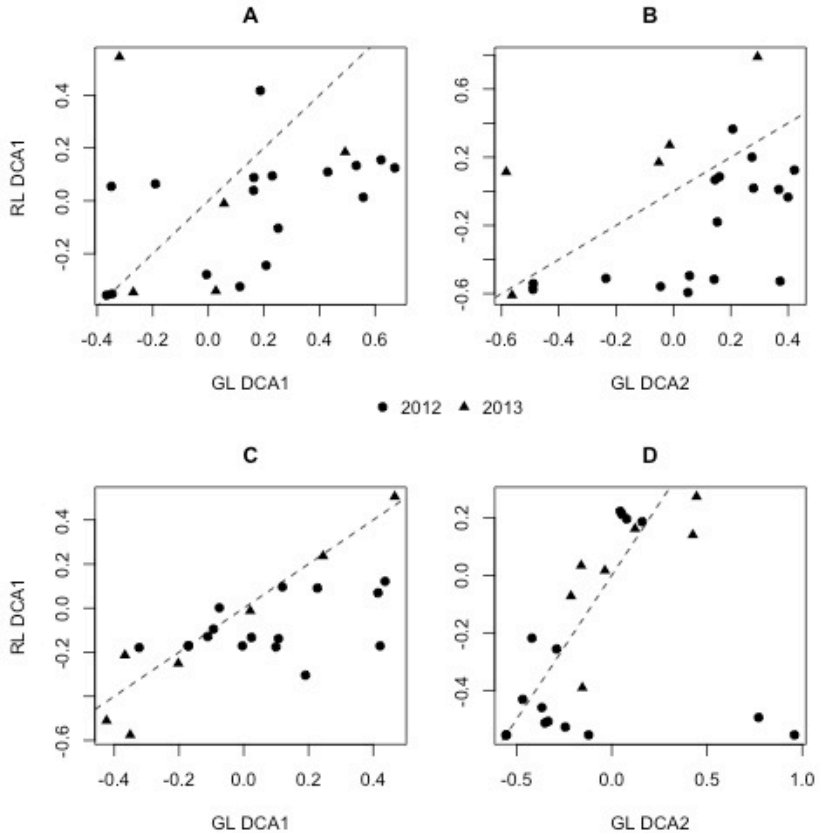


Figure A1.4 Plotting DCA axis scores from RL against those from GL for cyanobacterial (A and B) and cyanophage (C and D) populations. The dotted line represents a line of slope=1 with an intersect of (0,0). Points that fall on the dotted line are points at which GL and RL populations had a similar structure, while points falling away from the line show differing population structures. Interestingly, points generally fall near the 1:1 line for the cyanophage analysis, with the exception of DCA1 points that correspond to 2012 samples. On this axis, there is stability within the RL population, but variability within the GL population (C and D). When a similar comparison is done with the cyanobacterial DCA axis scores, few points fall on the 1:1 line (A and B). DCA was conducted through comparison of abundance of each of the cyanobacterial or cyanophage types at each timepoint. All time points were considered. For cyanobacteria, semi-quantitative abundances of each cyanobacterial type were used, for cyanophage, abundance of the cyanophage g20 gene was determined via qPCR. DCA was carried out in R using the vegan package.

Figure A1.5 Topographic image of GL and RL from google maps (maps.google.com). Location of sampling in GL and RL is indicated by the black triangles.

Table A1.1 Sequences used in the construction of Figure 2.5.

Tree Code	GI	GB	Citation	Location
C_FW1	29423524	AY224199.1	Crosbie et al 2003	European subalpine lake
H_Baltic1	145321650	EF513330.1	Haverkamp et al 2008	Baltic Sea
H_Baltic2	145321648	EF513328.1	Haverkamp et al 2008	Baltic Sea
C_FW2	24711738	AY151245.1	Crosbie et al 2003	European subalpine lake
C_FW3	146740642	AM709629.1	Crosbie et al 2003	European subalpine lake
E_FW1	18033220	AF330248.1	Ernst 2003	European subalpine lake
E_FW2	16588775	AF317075.1	Ernst 2003	European subalpine lake
H_Baltic3	145321610	EF513290.1	Haverkamp et al 2008	Baltic Sea
W_ETL1	222139167	FJ596210.1	Wu 2010	East Tibetan lakes
W_ETL2	222139168	FJ596211.1	Wu 2010	East Tibetan lakes
T_LC1	318086382	HQ687069.1	Tighe 2011	Lake Champlain
R_FW1	158905488	EU015872.1	Rueckert 2008	New Zeland lake
H_Baltic4	145321653	EF513333.1	Haverkamp et al 2008	Baltic Sea
E_FW3	18033219	AF330247.1	Ernst 2003	European subalpine lake
V_FW1	159139275	EU233405.1	Vareli 2009	Lake Ziros, Greece
W_ETL3	222139234	FJ596277.1	Wu 2010	East Tibetan lakes
W_ETL4	222139211	FJ596254.1	Wu 2010	East Tibetan lakes
W_ETL5	222139194	FJ596237.1	Wu 2010	East Tibetan lakes
H_Baltic5	169303531	EU386638.1	Haverkamp et al 2008	Baltic Sea
H_Baltic6	169303530	EU386637.1	Haverkamp et al 2008	Baltic Sea
Prochlorococcus marinus str. MIT 9303	18463020	AF397703.1	Rocap 2002	Marine
Prochlorococcus marinus str. MIT 9313	18463021	AF397704.1	Rocap 2002	Marine
Prochlorococcus marinus str. MIT 9211	18463019	AF397702.1	Rocap 2002	Marine
Synechococcus sp. WH 8101	18463045	AF397728.1	Rocap 2002	Marine
Synechococcus sp. MIT S9220	18463022	AF397705.1	Rocap 2002	Marine
Synechococcus sp. WH 7803	18463044	AF397727.1	Rocap 2002	Marine
Synechococcus sp. WH 5701	18463046	AF397729.1	Rocap 2002	Marine
Cyanobium gracile	18463047	AF397730.1	Rocap 2002	Marine
GL2816ITS11_Type4			This Study	Green/Round Lake
GL2821ITS5_Type4			This Study	Green/Round Lake
GL2821ITS10_Type4			This Study	Green/Round Lake
GL2816ITS7_Type4			This Study	Green/Round Lake
GL2821ITS3_Type4			This Study	Green/Round Lake
GL2821ITS7_Type4			This Study	Green/Round Lake
GL2816ITS1_Type3			This Study	Green/Round Lake
GLcITS13001A_Type3			This Study	Green/Round Lake
GLcITS13002A_Type2			This Study	Green/Round Lake
GL2821ITS4_Type2			This Study	Green/Round Lake
GLcITS12005A_Type2			This Study	Green/Round Lake
GL2816ITS10_Type2			This Study	Green/Round Lake
GLcITS12004A_Type1			This Study	Green/Round Lake
GL2816ITS4_Type1			This Study	Green/Round Lake
GL2816ITS12_Type1			This Study	Green/Round Lake
GLcITS5_Type1			This Study	Green/Round Lake

GL2816ITS5_Type1	This Study	Green/Round Lake
GL2821ITS9_Type1	This Study	Green/Round Lake
GL2816ITS2_Type1	This Study	Green/Round Lake
GL2821ITS12_Type1	This Study	Green/Round Lake
GL2821ITS11_Type1	This Study	Green/Round Lake
GL2816ITS9_Type1	This Study	Green/Round Lake
GL2821ITS2_Type1	This Study	Green/Round Lake
GL2816ITS8_Type1	This Study	Green/Round Lake
GLcITS13004A_Type1	This Study	Green/Round Lake
GLcITS13003A_Type1	This Study	Green/Round Lake
GLcITS12002A_Type1	This Study	Green/Round Lake

Appendix 2. Supplemental Information for Chapter 4.

Table A2.1 g20 clone library sequence information and colors associated with Figure 3.

Name	Core #	Depth (cm)	Varve Counts	# g20 clones	Color in Fig 3
SA64	A	13.75	61-67	21	
SA74	A	15.25	70-76	16	
SA88	A	17.75	84-91	30	
SB152	B	26.5	150-153	23	
Green Lake May 2012	Surface Waters			38	
Green Lake May 2013	Surface Waters			4	
Green Lake July 2013	Surface Waters			16	
Green Lake September 2013	Surface Waters			17	
Round Lake May 2012	Surface Waters			51	
Round Lake June 2012	Surface Waters			11	
Round Lake September 2012	Surface Waters			7	
IGR May 2012	Illumina Assembled Contigs from Green Lake and Round Lake May 2012 Surface Waters			9	
Total				243	
Color Strips					
IGRL Library	Amplicon library of g20 sequences from Green Lake and Round Lake, May 2012				
May2014 Library Recruitment	Purified Phage library from Green Lake Surface Waters collected May, 2014			14.8 M Reads	

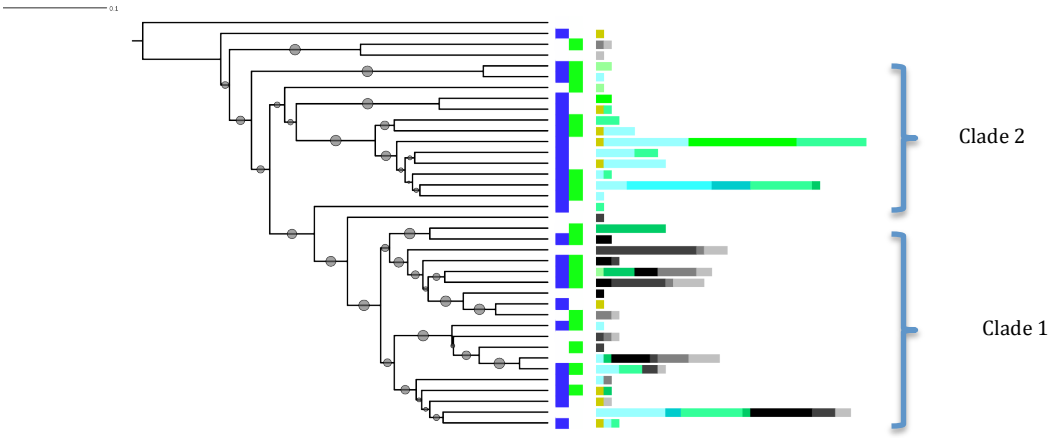


Figure A2.1 Phylogenetic tree of g20 clone library sequences clustered at 95% identity constructed using UPGMA. Circles indicate branches with bootstrap values >80, the blue inner strip indicates group representation within the IGR g20 library, the green strips indicate the representation within the May2014 library, and multicolored bars indicate the representation of different g20 clones to each cluster. Green colors are different time points in Green Lake, turquoise colors are different time points in Round Lake and blacks and greys indicate g20 sequences from various sediment layers (deep to surface: black to light grey). See Table A2.1 for a complete color key.

Figure A2.2 Approximate location of sediment cores sampling in GL. Map image taken from (Thompson et al., 1990).

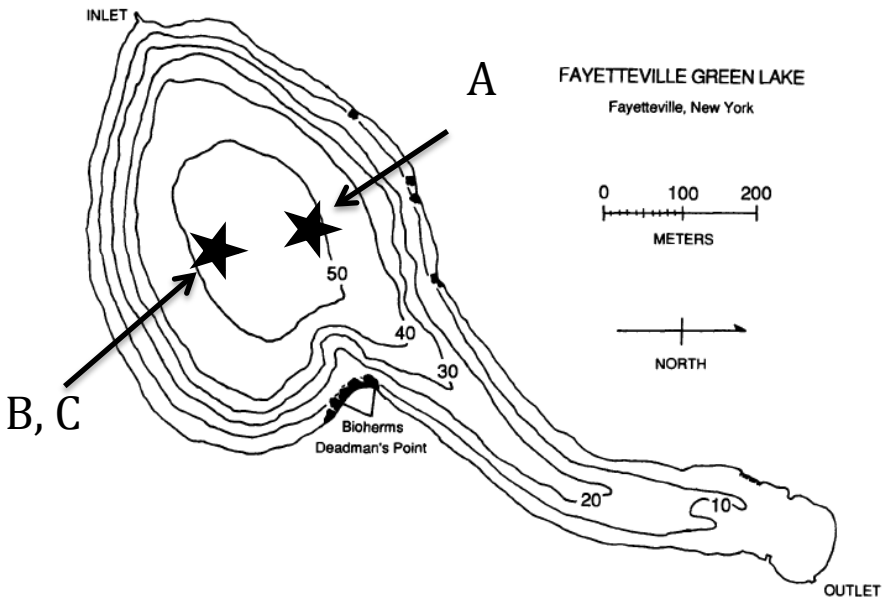


Figure A2.3 Digital X-radiographic images of sediment cores A, B and C and approximate location of sediment core fractions used for g20 clone library and viral metagenomic analyses.